METHODS FOR PRODUCTION OF OXYGENATED TERPENES

PRIORITY

This application claims the benefit of, and priority to, US Provisional Application No. 62/040,284 filed August 21, 2015, which is hereby incorporated by reference in its emtirety.

FIELD OF THE INVENTION

The present invention relates to oxygenated sesquiterpenes (e.g., nootkatone) and methods for their production and use. The invention also provides enzymes for the production of oxygenated sesquiterpenes (e.g., nootkatone) and methods for identifying, selecting, making and using these enzymes.

BACKGROUND OF THE INVENTION

The food and beverage industries as well as other industries such as the perfume, cosmetic and health care industries routinely use terpenes and/or terpenoid products as flavours and fragrances. By way of example, many sesquiterpene compounds are used in perfumery (e.g. patchoulol) and in the flavour industry (e.g., nootkatone) and many are extracted from plants. However, factors such as: (i) the availability and high price of the plant raw material; (ii) the relatively low terpene content in plant; and (iii) the tedious and inefficient extraction processes to produce sufficient quantities of terpene products on an industrial scale all have stimulated research on the biosynthesis of terpenes using plant-independent systems. Consequently, effort has been expended in developing technologies to engineer microorganisms for converting renewable resources such as glucose into terpenoid products. By comparison with traditional methods, microorganisms have the advantage of fast growth without the need for land to sustain development.

Many microorganisms use either the methylerythritol 4-phosphate (MEP) pathway or the melavonate (MVA) pathway to supply intermediates necessary to produce terpenoid products. These MEP or MVA pathways can include an endogenous

or an engineered MEP or MVA pathway or both. A detailed understanding of isoprenoid pathway engineering and optimization is disclosed in WO 2011/060057, US 2011/0189717, US 2012/107893, US 8,512,988 and Ajikumar et al (2010) *Science* 330 70-74, which discloses the production of various terpenoid compounds including sesquiterpene compounds such as nootkatone, which is an oxidised sesquiterpene produced from a valencene sesquiterpene substrate.

Nootkatone (4,4a,5,6,7,8-hexahydro-6-isopropenyl-4,4a-dimethyl-2(3II)-naphtalenone) is an important flavour constituent of grapefruit and is used commercially to flavour soft drinks and other beverages, as well as being used in perfumery. The conventional method for nootkatone preparation is by oxidation of valencene (see US 6,200,786 and US 8,097,442). The starting material valencene is expensive and thus methods that consume valencene are less commercially acceptable. Because of these drawbacks, there is a need for commercially feasible and sustainable methods to prepare nootkatone and associated products.

SUMMARY OF THE INVENTION

An object of the present invention is to provide sustainable production of oxygenated sesquiterpene products. Specifically, the present invention provides enzyme catalysts for the *ex vivo* or *in vivo* production of certain oxygenated sesquiterpenes. In some embodiments, the invention provides host cells engineered for the biosynthesis of the oxygenated sesquiterpenes. Another object of the present invention is to provide engineered cytochrome P450 (CYP450) enzymes for synthesis of oxygenated sesquiterpenes, including in some embodiments functional expression alongside a reductase counterpart in *E. coli*, yeast, or other host cell. The invention thereby harnesses the unique capability of this class of enzymes to conduct oxidative chemistry.

In one aspect, the invention provides a method for making an oxygenated product of a sesquiterpene. The method comprises contacting the sesquiterpene with *Stevia rebaudiana* Kaurene Oxidase (*Sr*KO) or derivative thereof having sesquiterpene oxidizing activity. Surprisingly, the wild type *Sr*KO enzyme was shown to have activity on a sesquiterpene substrate even though its natural activity is understood to act

on a diterpene substrate. Further, SrKO enzyme showed unique activities including oxygenation to the ketone, nootkatone), which requires two oxygenation cycles, and produced different oxygenated terpene products including hydroxygermacra-1(10)5-diene, and murolan-3,9(11) diene-10-peroxy. These activities are distinct from other P450 enzymes tested, which produced only one of the stereoisomers of the hydroxylated product (e.g., β-nootkatol), as the major product and/or produced only minor amounts of nootkatone.

In some embodiments, the method takes place in an *ex vivo* (e.g., cell free) system. In other embodiments, the sesquiterpene substrate and the *Sr*KO or derivative thereof are contacted in a cell expressing the *Sr*KO, such as a bacterium (e.g., *E. coli*). The oxygenated product of a sesquiterpene may be recovered, or may be the substrate for further chemical transformation. Functional expression of wild type cytochrome P450 in *E. coli* has inherent limitations attributable to the bacterial platforms (such as the absence of electron transfer machinery and cytochrome P450 reductases, and translational incompatibility of the membrane signal modules of P450 enzymes due to the lack of an endoplasmic reticulum). Thus, in some embodiments the *Sr*KO enzyme is modified for functional expression in an *E. coli* host cell, for example, by replacing a portion of the *Sr*KO N-terminal transmembrane region with a short peptide sequence that stabilizes interactions with the *E. coli* inner membrane and/or reduces cell stress.

In some embodiments, the *Sr*KO derivative has at least one mutation with respect to the wild type *Sr*KO that increases valencene oxidase activity (e.g., increases production of nootkatone). For example, the *Sr*KO may have from 1 to 50 mutations independently selected from substitutions, deletions, or insertions relative to wild type *Sr*KO (SEQ ID NO:37) or an *Sr*KO modified for expression and activity in *E. coli* (*e.g.,* SEQ ID NOS: 38_or 55). For example, the *Sr*KO derivative may have from 1 to 40 mutations, from 1 to 30 mutations, from 1 to 20 mutations, or from 1 to 10 mutations relative to *Sr*KO (SEQ ID NOS: 37, 38, or 55). In these or other embodiments, the *Sr*KO derivative may comprise an amino acid sequence having at least 50% sequence identity, or at least 60% sequence identity, or at least 70% sequence identity, or at least 80% sequence identity, or at least 90% sequence identity to *Sr*KO (SEQ ID NOS: 37, 38, or 55), and has valencene oxidase activity. The *Sr*KO in various embodiments

maintains valencene oxidase activity, or has increased valencene oxidase activity as compared to the wild type enzyme in an *ex vivo* or bacterial system (e.g. *E coli*). Various mutations of *Sr*KO which may maintain or enhance valencene oxidase activity are listed in **Table 2** and **Table 6**. Thus, in various embodiments, the *Sr*KO may have at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 mutations selected from **Table 2** and/or **Table 6**. Exemplary derivatives of *Sr*KO, also referred to herein as "valencene oxidase" or "VO" are represented by SEQ ID NOS: 104 and 105, which may further be derivatized for improvements in desired activity. Mutations may be selected empirically for increases in oxygenated sesquiterpene titer, or selected by *in silico* evaluation, or both.

In accordance with aspects of the invention, oxygenated sesquiterpene products are obtainable by contacting a sesquiterpene substrate with *Stevia rebaudiana* Kaurene Oxidase (*Sr*KO) or derivative thereof having valencene oxidizing activity. Unlike other CYP450 enzymes, when a *Sr*KO enzyme is used with valencene sesquiterpene substrate, it produces a different oxygenated terpene product profile that can include hydroxygermacra-1(10)5-diene, murolan-3,9(11) diene-10-peroxy, nootkatol, and nootkatone. By comparison, other CYP450's having the activity of hydroxylating valencene did not produce significant amounts of the ketone (nootkatone), which requires two oxygenation cycles. See **Table 4** and **Figure 7**.

In various embodiments, the sesquiterpene substrate is (or the predominant sesquiterpene substrate is) valencene, germacrene (A, B, C, D, or E), farnesene, farnesol, nootkatol, patchoulol, cadinene, cedrol, humulene, longifolene, and/or bergamotene, β-ylangene, β-santalol, β-santalene, α-santalene, α-santalol, ß-vetivone, a-vetivone, khusimol, bisabolene, β-aryophyllene, longifolene; α-sinensal; α-bisabolol, (-)-β-copaene, (-)-α-copaene, 4(Z),7(Z)-ecadienal, cedrol, cedrene, cedrol, guaiol, (-)-6,9-guaiadiene, bulnesol, guaiol, ledene, ledol, lindestrene, and alpha-bergamotene. In some embodiments, the predominant sesquiterpene substrate is valencene, and the predominant oxygenated product is nootkatone and/or nootkatol.

The invention, when applied *in vivo*, is applicable to a wide array of host cells. In some embodiments, the host cell is a microbial host, such as a bacterium selected

from *E. coli*, *Bacillus subtillus*, or *Pseudomonas putida*; or a yeast, such as a species of *Saccharomyces*, *Pichia*, or *Yarrowia*, including *Saccharomyces cerevisiae, Pichia pastoris,* and *Yarrowia lipolytica*.

In some embodiments, the host cell produces isopentyl pyrophosphate (IPP), which acts as a substrate for the synthesis of the sesquiterpene. In some embodiments, the IPP is produced by metabolic flux through an endogenous or heterologous methylerythritol phosphate (MEP) or mevalonic acid (MVA) pathway. In some embodiments, the sesquiterpene is produced at least in part by metabolic flux through an MEP pathway, and wherein the host cell has at least one additional copy of a *dxs*, *ispD*, *ispF*, and/or *idi* gene.

In some embodiments, the host cell expresses a farnesyl pyrophosphate synthase (FPPS), which produces farnesyl pyrophosphate (FPP) from IPP or DMAPP. The host cell may further express a heterologous sesquiterpene synthase to produce the desired sesquiterpene scaffold. For example, in some embodiments the cell expresses a valencene synthase. Several valencene synthase enzymes are known including *Vitis vinifera* valencene synthase (VvVS) (SEQ ID NO:1) or *Citrus sinensus* valencene synthase (CsVS) (SEQ ID NO: 12), which may be employed with the present invention, or alternatively a derivative of the VvVS or CsVS. Exemplary derivative VvVS enzymes are disclosed herein. In certain embodiments, the sesquiterpene synthase is a valencene synthase selected from Vv1M1 (SEQ ID NO:3), Vv2M1 (SEQ ID NO:5), Vv1M5 (SEQ ID NO:7), Vv2M5 (SEQ ID NO:9), or VS2 (SEQ ID NO:11), as disclosed herein.

The *Sr*KO or derivative thereof acts on the sesquiterpene (*e.g.*, valencene) to produce the oxygenated terpene product. In some embodiments the *Sr*KO is a fusion protein with a cytochrome P450 reductase partner (*e.g.*, *Sr*CPR), allowing the cofactor to be efficiently regenerated. In other embodiments, a P450 reductase is provided (e.g., to *in vitro* system) or expressed in the host cell separately, and may be expressed in the same operon as the *Sr*KO in some embodiments. In some embodiments, the CPR enzyme is expressed separately, and the gene may be integrated into the host cell genome in some embodiments. Various exemplary CPR enzymes are disclosed herein,

and which may be derivatized to improve oxygenated sesquiterpenoid titer and/or to improve P450 efficiency.

In some embodiments, the host cell expresses one or more enzymes that further direct oxygenated product to nootktone, such as the expression of one or more alcohol dehydrogenase (ADH) enzymes. Exemplary ADH enzymes are disclosed herein.

In other aspects, the invention provides a method for making a product containing an oxygenated sesquiterpene, which comprises incorporating the oxygenated sesquiterpene prepared and recovered according to the methods described herein into a consumer or industrial product. For example, the product may be a flavor product, a fragrance product, a cosmetic, a cleaning product, a detergent or soap, or a pest control product. In some embodiments, the oxygenated product recovered comprises nootkatone, and the product is a flavor product selected from a beverage, a chewing gum, a candy, or a flavor additive.

In other aspects, the invention provides engineered *Sr*KO enzymes having enhanced valencene oxidase activity as compared to wild type, as well as host cells producing an oxygenated sesquiterpene as described herein, and which express all of the enzyme components for producing the desired oxygenated sesquiterpene from isopentyl pyrophosphate (IPP). For example, the host cell in various embodiments expresses a farnesyl pyrophosphate synthase, a sesquiterpene synthase, and the *Sr*KO or derivative thereof. IPP may be produced through the MEP and/or MVA pathway, which may be endogenous to the host cell, and which may be enhanced through expression of heterologous enzymes or duplication of certain enzymes in the pathway. Host cells include various bacteria and yeast as described herein. The oxygenated sesquiterpene (e.g., nootkatone and/or nootkatol) may be recovered from the culture, and/or optionally may act as the substrate for further chemical transformation in the cell or *ex vivo* system.

In another aspect, the invention provides sesquiterpene-containing oil produced by the methods and host cells described herein. In some embodiments, the oil comprises hydroxygermacra-1(10)5-diene, murolan-3,9(11) diene-10-peroxy,

nootkatol, and nootkatone. In some embodiments, the predominant oxygenated products of valencene is nootkatone and/or nootkatol.

In another aspect, there is provided an *Sr*KO crystal model structure (CMS) based on the structural coordinates of P45017A1 (which catalyzes the biosynthesis of androgens). The CMS, including the terpene binding pocket domain (TBD) that comprises a terpene binding pocket (TBP) and a terpene (e.g., valencene) bound to the TBD, is illustrated in **Figures 8A** and **8B**. This *Sr*KO crystal model structure (CMS) facilitates *in-silico* testing of *Sr*KO derivatives. In part aided by this homology model, the present disclosure illustrates the use of several mutational strategies to identify increases or improvements in sesquiterpene oxygenation activity, including back-to-consensus mutagenesis, site-saturation mutagenesis, and recombination library screening.

Additional aspects and embodiments of the invention will be apparent from the following detailed description.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows a scheme for biosynthesis of valencene, which is a substrate for *Sr*KO in accordance with the present disclosure.

Figure 2 depicts the fold productivities for site-directed mutants made to VvVS. 46 of the 225 point mutations convey an average improvement in productivity of valencene of at least 20% compared to the wild-type WT VvVS. Figure 2 shows the number of VvVS mutants (y- axis) exhibiting certain levels of productivity (x-axis) versus the wild type.

Figure 3 (A and **B)** provides the amino acid and nucleotide sequences of valencene synthases. **Figure 3A** shows amino acid and nucleotide sequences from *Vitis vinifera* wild-type (WT) (VvVS) (SEQ ID NOS: 1 and 2) and derivatives Vv1M1 (SEQ ID NOS: 3 and 4), Vv2M1 (SEQ ID NOS: 5 and 6), Vv1M5 (SEQ ID NOS: 7 and 8), Vv2M5 (SEQ ID NOS: 9 and 10), and amino acid sequence of the derivative VS2 (SEQ ID NO:11); as well as amino acid sequence for *Citrus sinensus* wild-type

(CsVS) (SEQ ID NO:12). **Figure 3B** shows an alignment of wild-type VvVS and CsVS sequences, and the engineered Vv2M5 and VS2 sequences.

Figure 4 (A and **B)** provides the amino acid and nucleotide sequences of various CYP450 (Cytochrome P450) enzymes having activity on sesquiterpene scaffolds. **Figure 4A** shows sequences of wild type amino acid sequences and amino acid and nucleotide sequences engineered for bacterial expression: ZzHO (SEQ ID NO: 13, 14, and 15 respectively), BsGAO (SEQ ID NO: 16, 17, and 18, respectively), HmPO (SEQ ID NO: 19, 20, and 21 respectively), LsGAO (SEQ ID NO: 22, 23, and 24, respectively), NtEAO (SEQ ID NO: 25, 26, and 27, respectively), CpVO (SEQ ID NO: 28, 29, and 30, respectively), AaAO (SEQ ID NO: 31, 32, and 33, respectively), AtKO (SEQ ID NO: 34, 35, and 36 respectively), SrKO (SEQ ID NO: 37, 38, and 39 respectively), PpKO (SEQ ID NO: 40, 41, and 42, respectively), BmVO (SEQ ID NO: 43 and SEQ ID NO: 44, respectively), PsVO (SEQ ID NO: 45 and SEQ ID NO: 46, respectively), PoLO (SEQ ID NO: 47 and SEQ ID NO: 48, respectively), CiVO (SEQ ID NO: 49, 50, and 51 respectively), HaGAO (SEQ ID NO: 52, 53, and 54, respectively). **Figure 4B** shows amino acid sequences of engineered Valencene Oxidase enzymes based on the SrKO scaffold (SEQ ID NOS: 55-61).

Figures 5A and **5B** depict construct designs for expression of MEP, terpene and terpenoid synthases, and P450 enzymes in *E. coli*. **Figure 5A** shows strain configuration of upstream MEP pathway genes and the two plasmids harboring downstream pathway genes. **Figure 5B** shows construction of P450 fusions, whereby N-terminal regions of both the P450 and CPR (Cytochrome P450 reductase) are truncated and an exemplary leader sequence (MALLLAVF -- SEQ ID NO:112) (8RP) is added while the two are fused with a short linker peptide.

Figure 6 (A-D) provides the amino acid and nucleotide sequences of various CPR (Cytochrome P450 reductase) enzymes with sequence alignments. In **Figure 6A**: *Stevia rebaudiana* (*Sr*)CPR (SEQ ID NOS: 62 and 63), *Stevia rebaudiana* (*Sr*)CPR1 (SEQ ID NOS: 76 and 77), *Arabidopsis thaliana* (*At*)CPR (SEQ ID NOS: 64 and 65), *Taxus cuspidata* (*Tc*) CPR (SEQ ID NOS: 66 and 67), *Artemisia annua* (*Aa*)CPR (SEQ ID NOS: 68 and 69), *Arabidopsis thaliana* (*At*)CPR1 (SEQ ID NOS: 70 and 71),

Arabidopsis thaliana (*At*)CPR2 (SEQ ID NOS: 72 and 73), *Arabidopsis thaliana* (*At*)R2 (SEQ ID NOS: 74 and 75); *Stevia rebaudiana* (*Sr*)CPR2 (SEQ ID NOS: 78 and 79); *Stevia rebaudiana* (*Sr*)CPR3 (SEQ ID NOS: 80 and 81); *Pelargonium graveolens* (*Pg*)CPR (SEQ ID NO: 82 and 83). **Figure 6B** shows an alignment of amino acid sequences for *Arabidopsis thaliana* and *Artemisia annua* CPR sequences (SEQ ID NOS:72, 74, 68, 64, and 70). **Figure 6C** shows an alignment of *Stevia rebaudiana* CPR sequences (SEQ ID NOS: 78, 80, 62, and 76). **Figure 6D** shows an alignment of eight CPR amino acid sequences (SEQ ID NO: 74, 72, 82, 68, 80, 62, 78, and 76).

Figure 7 provides GC-chromatographs which show the different activities of various CYP450 enzymes, as expressed in valencene-producing *E. coli* along with CPR partners as described in Example 2. Strains were cultured for four days and extracted with Methyl Tert-Butyl Ether (MTBE). 1 µl of MTBE was injected through GC-MS and the product profiles were monitored by comparing with a MS library. From top to bottom: Taxus 5-alpha hydroxylase, Cichorium intybus (CiVO) P450 (SEQ ID NO:50), Hyoscyamus muticus (HmPO) P450 (SEQ ID NO:20), and *Sr*KO (SEQ ID NO:38).

Figures 8A and **8B** illustrate a homology model of *Sr*KO and its active site. The *Sr*KO homology model is based on the known mutant P45017A1 (the crystal structure of membrane-bound cytochrome P450 17 A1 as disclosed in DeVore NM and Scott EE (*Nature*, 482, 116-119, 2012), which catalyzes the biosynthesis of androgens in human. The position of the heme is shown as sticks. **Figure 8B** depicts a structural model of *Sr*KO active site with valencene docked in its α-binding mode. Secondary structure motifs (B-C Loop and I-Helix) and amino acids targeted for mutagenesis are shown.

Figure 9 shows optimizing the valencene oxidase (VO) N-terminal membrane anchor. The N-terminus of *E. coli* yhcB was selected as a membrane anchor sequence, which provides a single-pass transmembrane helix. The length of the anchor (from 20 to 24 amino acids) and the VO N-terminal truncation length (from 28 to 32 amino acids) were screened for improvements in oxygenation titer.

Figure 10 shows that a truncation length of 29, and a 20 amino acid N-terminal anchor based on *E. coli* yhcB, led to a 1.2-fold increase in total oxygenated titer compared to the average of controls.

Figure 11 illustrates an exemplary downstream pathway for expression in the host cell, for conversion of farnesyl diphosphate to nootkatone. Farnesyl diphosphate (produced from IPP/DMAPP by an expressed Farnesyl Pyrophosphate Synthase) is converted to valencene by the action of a Valencene Synthase (VS), which is oxidized by a Valencene Oxidase (VO), such as SrKO or an engineered derivative described herein. The VO cofactor is regenerated by a cytochrome P450 reductase (CPR). The products of oxidation by VO can include nootkatol and nookatone, which can be further directed to nootkatone by the action of an Alcohol Dehydrogenase (ADH).

Figure 12 shows the oxygenation profile for a strain expressing VO1-L-SrCPR. The oxygenation profile includes the single oxygenation products of β-nootkatol and α-nootkatol along with the two-step oxygenation product, nootkatone.

Figure 13 (**A** and **B**) shows evaluation of mutations identified using a back-to-consensus strategy in wild-type *Sr*KO, translated into an engineered valencene oxidase background (n22yhcB_t30VO1**)**. More than 50% of the mutations resulted in a 1.2 to 1.45 fold improvement in total oxygenated titers. Panel (**A**) shows titer in mg/L. Panel (**B**) shows fold change in oxygenated titer.

Figure 14 shows results of secondary screening of back-to-consensus mutations, N-terminal anchor optimization, and site-saturation mutagenesis (SSM). Several mutations were identified that show a 1.1 to 1.4-fold improvement in oxygenated titers.

Figure 15 shows performance of select VO1 variants at 33°C. Six mutations were identified that maintained improved productivities at 33°C.

Figure 16 (**A** and **B**) shows results from primary screening of the recombination library. Several variants (shown) exhibited up to 1.35-fold improvement in oxygenated product titer. There was a shift in profile to more (+)-nootkatone and

higher oxygenation capacity for select variants. Panel (**A**) shows oxygenated product in mg/L. Panel (**B**) plots the fold change in oxygenation capacity (nootkatols require only one oxygenation cycle from valencene, while nootkatone requires two oxygenation cycles).

Figure 17 shows oxygenation capacity at 34°C and 37°C for select VO recombination library variants.

Figure 18 shows oxygenation titer at 34°C and 37°C after re-screen of lead VO variants. C6(1) (R76K, M94V, T131Q, I390L, T468I) had the highest oxygenation capacity at 37°C, and was designated VO2.

Figure 19 shows screening of cytochrome P450 reductase (CPR) orthologs for enhanced valencene oxidase activity (30°C). *Sr*CPR3 shows increased oxygenation titer and higher production of Nootkatone.

Figure 20 shows screening of CPR orthologs at 34°C. *Sr*CPR3 and *Aa*CPR exhibit ~1.3-fold improvement in oxygenated titer, even at higher the higher temperature.

Figure 21 shows conversion of nootkatols to nootkatone with an alcohol dehydrogenase. Four ADH orthologs (vvDH, csABA2, bdDH, and zzSDR) were identified that convert nootkatol to (+)-nootkatone, resulting in a 3-fold increase in (+)-nootkatone titers.

Figure 22 (**A** and **B**) depicts alcohol dehydrogenase enzymes. **Figure 22A** shows amino acid and nucleotide sequences including those for *Rhodococcus erythropolis* (*Re*)CDH (SEQ ID NOS: 84 and 85), *Citrus sinensus* (*Cs*)DH (SEQ ID NOS: 86 and 87), *Citrus sinensus* (*Cs*)DH1 (SEQ ID NOS: 88 and 89), *Citrus sinensus* (*Cs*)DH2 (SEQ ID NOS: 90 and 91), *Citrus sinensus* (*Cs*)DH3 (SEQ ID NOS: 92 and 93), *Vitis vinifera* (*Vv*)DH (SEQ ID NOS: 94 and 95), *Vitis vinifera* (*Vv*)DH1 (SEQ ID NOS: 96 and 97, *Citrus sinensus* (*Cs*)ABA2 (SEQ ID NOS: 98 and 99), *Brachypodium distachyon* (*Bd*)DH (SEQ ID NO: 100 and 101), *Zingiber zerumbet* (*Zz*)SDR (SEQ ID NOS: 102 and 103). **Figure 22B** shows an alignment of the amino acid sequences.

Figure 23 (**A** and **B**) shows alignments of several engineered valencene oxidase (VO) variants. In **Figure 23A**: 8rp-t20SrKO (SEQ ID NO: 106) is the *Sr*KO sequence with a 20-amino acid truncation at the N-terminus, and the addition of an 8-amino acid membrane anchor. 8rp-t20VO0 (SEQ ID NO: 107) has a truncation of 20 amino acids of the *Sr*KO N-terminus, the addition of an 8-amino acid N-terminal anchor, and a single mutation at position 499 (numbered according to wild-type *Sr*KO). n22yhcB-t30VO1 (SEQ ID NO: 104) has a 30-amino acid truncation of the *Sr*KO N-terminus, a membrane anchor based on 22 amino acids from *E. coli* yhcB, and eight point mutations at positions 46, 231, 284, 383, 400, 444, 488, and 499 (with respect to *Sr*KO wild-type). n22yhcB-t30VO2 (SEQ ID NO: 105) has a 30-amino acid truncation of the *Sr*KO N-terminus, a membrane anchor based on 22 amino acids from *E. coli* yhcB, and nine point mutations at positions 76, 94, 131, 231, 284, 383, 390, 468, and 499 (with respect to *Sr*KO wild-type). **Figure 23B**, point mutations in VO0 (SEQ ID NO: 109), VO1 (SEQ ID NO: 110), and VO2 (SEQ ID NO: 111) are shown against wild-type *Sr*KO (SEQ ID NO: 108) (all shown with the wild-type *Sr*KO N-terminus for convenience).

DETAILED DESCRIPTION OF THE INVENTION

The present invention in various aspects provides methods for making oxygenated terpenes or terpenoids in *ex vivo* or in cell systems. The invention further provides engineered or modified enzymes, polynucleotides, and host cells for use in such methods. The invention in various embodiments is directed to a method to produce nootkatone using an *Sr*KO enzyme. Surprisingly, it was found that the *Sr*KO enzyme can be used to catalyze sesquiterpene oxidation (e.g., valencene oxidation to nootkatol and nootkatone).

As used herein, *Sr*KO refers to ent-kaurene oxidase CYP701A5 [Stevia rebaudiana] with Accession No AAQ63464.1 (SEQ ID NO:37). *Sr*KO and its activity on diterpenes (and kaurene in particular) are known and are described in, for example, US 2012/0164678, which is hereby incorporated by reference in its entirety. It is a member of the CYP70 family of cytochrome p450 enzymes (CYP450). An *Sr*KO

sequence modified for expression in *E. coli* is shown as SEQ ID NO: 38. As shown herein, *Sr*KO is active on sesquiterpene substrates (e.g., valencene), producing nootkatol and nootkatone, which are valuable terpenoid compounds. These oxygenation activities and product profiles (e.g., increasing production of nootkatone) can be further refined by mutagenesis of the *Sr*KO using processes (and aided by *in silico* models) described in detail herein.

As used herein, the term "*Sr*KO derivative" or "engineered *Sr*KO" refers to an amino acid sequence that has substantial structural and/or sequence identity with *Sr*KO, and catalyzes oxygenation of a sesquiterpene scaffold, such as valencene. *Sr*KO enzymes engineered for the oxygentation of valencene are also referred to herein as "valencene oxidase" or "VO" enzymes. Generally, derivatives comprise mutated forms of *Sr*KO having at least one mutation that increases the activity of the enzyme for the valencene substrate or for the production of nootkatone and/or other products. Some *Sr*KO mutations are provided in **Table 2**. Some such additional *Sr*KO mutations are provided in **Table 6**.

The term "contacting" means that the components are physically brought together, whether *in vivo* through co-expression of relevant protein products (e.g., sesquiterpene synthase and CYP450) in a host cell, or by adding or feeding a substrate of interest to a host cell expressing an *Sr*KO or derivative thereof, or *in vitro* (or "*ex vivo*") by adding sesquiterpene substrate to purified P450 enzyme or cellular extract or partially purified extract containing the same. The terms *in vitro* and *ex vivo* refer to a cell free system, and may be performed in a reaction tube or well.

As used herein, "terpenes" are a large and varied class of hydrocarbons that have a simple unifying feature, despite their structural diversity. According to the "isoprene rule", all terpenes consist of isoprene (C5) units. This fact is used for a rational classification depending on the number of such units. Monoterpenes comprise 2 isoprene units and are classified as (C10) terpenes, sesquiterpenes comprise 3 isoprene units and are classified as (C15) terpenes, diterpenes comprise 4 isoprene units and are classified as (C20) terpenes, sesterterpenes (C25), triterpenes (C30) and rubber (C5)n. They occur as acyclic or mono- to pentacyclic derivatives with alcohol, ether,

ester, aldehyde, or ketone groups (the so called "terpenoids"), everywhere in organisms, particularly in higher plants, and are characteristic of the individual type of plants. Terpenes such as Monoterpenes (C10), Sesquiterpenes (C15) and Diterpenes (C20) are derived from the prenyl diphosphate substrates, geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) respectively through the action of a very large group of enzymes called the terpene (terpenoid) synthases. These enzymes are often referred to as terpene cyclases since the product of the reactions are cyclised to various monoterpene, sesquiterpene and diterpene carbon skeleton products. Many of the resulting carbon skeletons undergo subsequence oxygenation by cytochrome p450 hydrolysase enzymes to give rise to large families of derivatives. The technical syntheses of top-selling flavours and fragrances can start from terpenes which can also serve as excellent solvents or diluting agents for dyes and varnishes. Natural or synthetic resins of terpenes are used and also many pharmaceutical syntheses of vitamins and insecticides start from terpenes. As used herein, the term "terpene" or "sesquiterpene" (for example) includes corresponding terpenoid or sesquiterpenoid compounds.

As used herein, the term "oxygenated sesquiterpene" refers to a sesquiterpene scaffold having one or more oxygenation events, producing a corresponding alcohol, aldehyde, carboxylic acid and/or ketone.

As used herein, the term "MEP pathway" refers to the (2-C-methyl-D-erythritol 4-phosphate) pathway, also called the MEP/DOXP (2-C-methyl-D-erythritol 4-phosphate/l-deoxy-D-xylulose 5-phosphate) pathway or the non-mevalonate pathway or the mevalonic acid-independent pathway. In the MEP pathway, pyruvate and D-glyceraldehyde-3-phosphate are converted via a series of reactions to IPP and DMAPP. The pathway typically involves action of the following enzymes: 1-deoxy-D-xylulose-5-phosphate synthase (Dxs), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (IspC), 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE), 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF), 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (IspG), and isopentenyl diphosphate isomerase (IspH). The MEP pathway, and the genes and

enzymes that make up the MEP pathway, are described in US 8,512,988, which is hereby incorporated by reference in its entirety. For example, genes that make up the MEP pathway include *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, and *ispA*.

As used herein, the MVA pathway refers to the biosynthetic pathway that converts acetyl-CoA to IPP. The mevalonate pathway typically comprises enzymes that catalyze the following steps: (a) condensing two molecules of acetyl-CoA to acetoacetyl-CoA (*e.g.*, by action of acetoacetyl-CoA thiolase); (b) condensing acetoacetyl-CoA with acetyl-CoA to form hydroxymethylglutaryl-CoenzymeA (HMG-CoA) (*e.g.*, by action of HMG-CoA synthase (HMGS)); (c) converting HMG-CoA to mevalonate (*e.g.*, by action of HMG-CoA reductase (HMGR)); (d) phosphorylating mevalonate to mevalonate 5-phosphate (*e.g.*, by action of mevalonate kinase (MK)); (e) converting mevalonate 5-phosphate to mevalonate 5-pyrophosphate (*e.g.*, by action of phosphomevalonate kinase (PMK)); and (f) converting mevalonate 5-pyrophosphate to isopentenyl pyrophosphate (*e.g.*, by action of mevalonate pyrophosphate decarboxylase (MPD)). The MVA pathway, and the genes and enzymes that make up the MEP pathway, are described in US 7,667,017, which is hereby incorporated by reference in its entirety.

As used herein, the term "cytochrome P450 reductase partner" or "CPR partner" refers to a cytochrome P450 reductase capable of regenerating the cofactor component of the cytochrome P450 oxidase of interest (e.g., *Sr*KO) for oxidative chemistry. For example, *Sr*CPR is a natural CPR partner for *Sr*KO. In some embodiments, the CPR partner is not the natural CPR partner for *Sr*KO. In some embodiments employing *in vivo* production of oxygenated sesquiterpene, the *Sr*KO and *Sr*CPR are co-expressed as separate proteins, or in some embodiments are expressed as a fusion protein.

The similarity of nucleotide and amino acid sequences, i.e. the percentage of sequence identity, can be determined via sequence alignments. Such alignments can be carried out with several art-known algorithms, such as with the mathematical algorithm of Karlin and Altschul (Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5877), with hmmalign (HMMER package, http://hmmer.wustl.edu/) or with the CLUSTAL algorithm (Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) Nucleic

Acids Res. 22, 4673-80). The grade of sequence identity (sequence matching) may be calculated using e.g. BLAST, BLAT or BlastZ (or BlastX). A similar algorithm is incorporated into the BLASTN and BLASTP programs of Altschul et al (1990) J. Mol. Biol. 215: 403-410. BLAST polynucleotide searches can be performed with the BLASTN program, score = 100, word length = 12.

BLAST protein searches may be performed with the BLASTP program, score = 50, word length = 3. To obtain gapped alignments for comparative purposes, Gapped BLAST is utilized as described in Altschul et al (1997) Nucleic Acids Res. 25: 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs are used. Sequence matching analysis may be supplemented by established homology mapping techniques like Shuffle-LAGAN (Brudno M., Bioinformatics 2003b, 19 Suppl 1:154-162) or Markov random fields.

"Conservative substitutions" may be made, for instance, on the basis of similarity in polarity, charge, size, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the amino acid residues involved. The 20 naturally occurring amino acids can be grouped into the following six standard amino acid groups:

(1) hydrophobic: Met, Ala, Val, Leu, Ile;

(2) neutral hydrophilic: Cys, Ser, Thr; Asn, Gin;

(3) acidic: Asp, Glu;

(4) basic: His, Lys, Arg;

(5) residues that influence chain orientation: Gly, Pro; and

(6) aromatic: Trp, Tyr, Phe.

As used herein, "conservative substitutions" are defined as exchanges of an amino acid by another amino acid listed within the same group of the six standard amino acid groups shown above. For example, the exchange of Asp by Glu retains one negative charge in the so modified polypeptide. In addition, glycine and proline may

be substituted for one another based on their ability to disrupt a-helices. Some preferred conservative substitutions within the above six groups are exchanges within the following sub-groups: (i) Ala, Val, Leu and He; (ii) Ser and Thr; (ii) Asn and Gin; (iv) Lys and Arg; and (v) Tyr and Phe.

As used herein, "non-conservative substitutions" or "non-conservative amino acid exchanges" are defined as exchanges of an amino acid by another amino acid listed in a different group of the six standard amino acid groups (1) to (6) shown above.

In one aspect, the invention provides a method for making an oxygenated product of a sesquiterpene. In various embodiments, the sesquiterpene substrate is (or the predominant sesquiterpene substrate is) valencene, germacrene (A, B, C, D, or E), farnesene, farnesol, nootkatol, patchoulol, cadinene, cedrol, humulene, longifolene, and/or bergamotene, β-ylangene, β-santalol, β-santalene, α-santalene, α-santalol, ß-vetivone, a-vetivone, khusimol, bisabolene, β-aryophyllene, Longifolene; α-sinensal; α-bisabolol, (-)-β-copaene, (-)-α-copaene, 4(Z),7(Z)-ecadienal, cedrol, cedrene, cedrol, guaiol, (-)-6,9-guaiadiene, bulnesol, guaiol, ledene, ledol, lindestrene, and alpha-bergamotene. In some embodiments, the predominant sesquiterpene substrate is valencene, and the predominant oxygenated product is nootkatone and/or nootkatol. In this context, the term “predominant” means that the particular sesquiterpene is present at a level higher than all other terpene or terpenoid species individually. In some embodiments, the predominant sesquiterpene (either the substrate or the oxygenated product after the reaction) makes up at least 25%, at least 40%, at least 50%, or at least 75% of the terpene or terpenoid component of the composition. In various embodiments involving *in vivo* production of oxygenated sesquiterpenes, the oxygenation product is recovered from the culture media, and can be fractionated to isolate or enrich for various components of the product, such as nootkatone. In some embodiments, Nootkatone is isolated and/or enriched, such that it makes up at least about 25%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, or at least about 90% of the sesquiterpene component (by weight), or makes up at least about 25%, at least about 40%, at least about 50%, at least

about 60%, at least about 70%, at least about 80%, at least about 90%, or about 100% of the collective amount of nootkatol and nootkatone (by weight).

In various embodiments, the invention comprises contacting a sesquiterpene with a terpene oxidizing P450 enzyme, or derivative thereof. The contacting may take place in a host cell or in a cell free system. The substrate for oxidation (e.g., the sesquiterpene), may be produced by the cells (e.g., through metabolic flux through the MEP or MVA pathways), or alternatively fed to the host cells expressing the P450 enzyme. The oxygenated product may be recovered, or be the substrate for further chemical transformation either in the cellular system or cell free system. **Table 1** below provides a list of exemplary P450 enzymes. While in certain embodiments the invention involves the use of the following P450 enzymes (optionally engineered to increase the oxygenation of valencene to nootkatone and/or nootkatol), a preferred enzyme in accordance with this disclosure is *Sr*KO. Exemplary oxygenated sesquiterpene products obtained by these reactions in accordance with the disclosure are shown in **Table 4**.

Table 1

#	**Species**	**Name**	**Native Substrate**	**Native Reaction Product**
1	*Zingiber zerumbet*	*zzHO*	α-humulene	8-hydroxy-α-humulene
2	*Barnadesia spinosa*	*BsGAO*	germacrene A	germacra-1(10),4,11(13)-trien-12-ol
3	*Hyoscyamus muticus*	*HmPO*	premnaspirodiene	solavetivol
4	*Latuca spicata*	*LsGAO*	germacrene A	germacra-1(10),4,11(13)-trien-12-ol
5	*Nicotiana tabacum*	*NtEAO*	5-epi-aristolochene	capsidiol
6	*Citrus x paradisi*	*CpVO*	valencene	nootkatol
7	*Artemesia annua*	*AaAO*	amorphadiene	artemisinic acid
8	*Arabidopsis thaliana*	*AtKO*	kaurene	kaurenoic acid
9	*Stevia rebaudiana*	*SrKO*	kaurene	kaurenoic acid
10	*Pseudomonas putida*	*PpKO*	kaurene	kaurenoic acid

11	*Bacillus megaterium*	*BmVO*	fatty acids	hydroxylated FAs
12	*Pleurotus sapidus*	*PsVO*	valencene	nootkatone
13	*Pleurotus ostreatus*	*PoLO*	unknown	unknown
14	*Cichorium intybus*	*CiVO*	valencene	nootkatone
15	*Helianthus annuus*	*HaGAO*	germacrene A	germacrene A acid

In various embodiments, the method comprises contacting the sesquiterpene with a protein comprising *Stevia rebaudiana* Kaurene Oxidase (*Sr*KO) or derivative thereof. In some embodiments the *Sr*KO is expressed in a host cell as described below, or is provided in a cell free system. For example, certain *in vitro* and *in vivo* systems for oxidizing terpenes with P450 enzymes are disclosed in US 7,211,420, which are hereby incorporated by reference. McDougle DR, Palaria A, Magnetta E, Meling DD, Das A. Functional Studies of N-terminally modified CYP2J2 epoxygenase in Model Lipid Bilayers, Protein Sci. 2013 22:964-79; Luthra, A., Gregory, M., Grinkova, Y. V., Denisov, I. G., Sligar, S. G. (2013) "Nanodiscs in the studies of membrane-bound cytochrome P450 enzymes." Methods Mol. Biol., 987, 115-127).

In some embodiments, the *Sr*KO derivative comprises an amino acid sequence that has from about 1 to about 50 mutations independently selected from substitutions, deletions, or insertions relative to *Sr*KO (SEQ ID NO: 37), or relative to an *Sr*KO enzyme modified at its N-terminus for functional expression in *E. coli* (SEQ ID NO:38 or 55) In various embodiments, the mutation or combination of mutations enhances the activity of the enzyme for oxygenation of valencene, such as the production of nootkatone. Protein modeling as described herein may be used to guide such substitutions, deletions, or insertions in the *Sr*KO sequence. For example, a structural model of the *Sr*KO amino acid sequence may be created using the coordinates for P45017A1. As demonstrated herein, such a homology model is useful for directing improvement of *Sr*KO for valencene oxygenation. Thus, in various embodiments, the *Sr*KO derivative may have from about 1 to about 45 mutations, about 1 to about 40 mutations, about 1 to about 35 mutations, from about 1 to about 30 mutations, about 1 to about 25 mutations, from about 1 to about 20 mutations, about 1 to about 15

mutations, about 1 to about 10 mutations, or from about 1 to about 5 mutations relative to *Sr*KO (SEQ ID NOS: 37, 38, or 55). In various embodiments, the *Sr*KO comprises a sequence having at least 5 or at least 10 mutations with respect to SEQ ID NO: 37, 38, or 55 but not more than about 20 or 30 mutations. In various embodiments, the *Sr*KO derivative may have about 1 mutation, about 2 mutations, about 3 mutations, about 4 mutations, about 5 mutations, about 6 mutations, about 7 mutations, about 8 mutations, about 9 mutations, about 10 mutations, about 11 mutations, about 12 mutations, about 13 mutations, about 14 mutations, about 15 mutations, about 16 mutations, about 17 mutations, about 18 mutations, about 19 mutations, about 20 mutations, about 21 mutations, about 22 mutations, about 23 mutations, about 24 mutations, about 25 mutations, about 26 mutations, about 27 mutations, about 28 mutations, about 29 mutations, about 30 mutations, about 31 mutations, about 32 mutations, about 33 mutations, about 34 mutations, about 35 mutations, about 36 mutations, about 37 mutations, about 38 mutations, about 39 mutations, about 40 mutations, about 41 mutations, about 42 mutations, about 43 mutations, about 44 mutations, about 45 mutations, about 46 mutations, about 47 mutations, about 48 mutations, about 49 mutations, or about 50 mutations relative to *Sr*KO (SEQ ID NO: 37,-38, or 55). SEQ ID NOS:37, and other WT enzymes disclosed herein, can optionally contain an Ala at position 2 where not present in the wild-type.

In these or other embodiments, the *Sr*KO derivative may comprise an amino acid sequence having at least about 50% sequence identity, at least about 55% sequence identity, at least about 60% sequence identity, at least about 65% sequence identity, at least about 70% sequence identity, at least about 75% sequence identity, at least about 80% sequence identity, at least about 85% sequence identity, or at least 90% sequence identity, or at least 91% sequence identity, or at least 92% sequence identity, or at least 93% sequence identity, or at least 94% sequence identity, or at least 95% sequence identity, or at least 96% sequence identity, or at least 97% sequence identity, or at least 98% sequence identity, or at least 99% sequence identity, to *Sr*KO (SEQ ID NO: 37, 38, or 55). In various embodiments, the *Sr*KO derivative has higher activity for the oxygenation of valencene than the wild type enzyme, such as a higher production of oxygenated oil upon contact with valencene substrate than the wild type enzyme (SEQ

ID NO:37) or the wild type enzyme as modified for functional expression in *E. coli*. For example, the *Sr*KO derivative may comprise an amino acid sequence having at least: about 50% identity, about 51% identity, about 52% identity, about 53% identity, about 54% identity, about 55% identity, about 56% identity, about 57% identity, about 58% identity, about 59% identity, about 60% identity, about 61% identity, about 62% identity, about 63% identity, about 64% identity, about 65% identity, about 66% identity, about 67% identity, about 68% identity, about 69% identity, about 70% identity, about 71% identity, about 72% identity, about 73% identity, about 74% identity, about 75% identity, about 76% identity, about 77% identity, about 78% identity, about 79% identity, about 80% identity, about 81% identity, about 82% identity, about 83% identity, about 84% identity, about 85% identity, about 86% identity, about 87% identity, about 88% identity, about 89% identity, about 90% identity, about 91% sequence identity, about 92% sequence identity, about 93% sequence identity, about 94% sequence identity, about 95% sequence identity, about 96% sequence identity, about 97% sequence identity, about 98% sequence identity, or about 99% sequence identity to *Sr*KO (SEQ ID NO: 37, 38, or 55).

In some embodiments, mutants are selected for an increase in production of oxygenated valencene, such as nootkatone. For example, the *Sr*KO derivative may have one, two, three, four or more mutations at positions selected from 46, 76, 94, 131, 231, 284, 383, 390, 400, 444, 468, 488 and 499, relative to SEQ ID NO:37. For example, in some embodiments the *Sr*KO is a derivative comprising an amino acid sequence having one or more (e.g, 2, 3, 4, or all) of the mutations selected from R76K, M94V, T131Q, F231L, H284Q, R383K, I390L, T468I, and T499N relative to SEQ ID NO:37. In some embodiments, the *Sr*KO derivative comprises an amino acid sequence selected from SEQ ID NOS:55-61, 104, or 105, which were engineered according to this disclosure to improve activity for oxygenation of valencene (e.g., production of nootkatone). In some embodiments, the derivative comprises an amino acid sequence having from one to twenty, or from one to ten, or from one to five mutations relative to a sequence selected from SEQ ID NOS: 55-61, 104, and 105 with the proviso that the amino acid sequence has one or more mutations at positions selected from 46, 76, 94, 131, 231, 284, 383, 390, 400, 444, 468, 488 and 499 relative to SEQ ID NO:37, or the

proviso that the *Sr*KO derivative comprises an amino acid sequence having one, two, three or more (or all) of the mutations selected from R76K, M94V, T131Q, F231L, H284Q, R383K, I390L, T468I, and T499N relative to SEQ ID NO:37.

In some embodiments, the invention provides a recombinant polynucleotide encoding the *Sr*KO derivative described above, which may be inserted into expression vectors for expression and optional purification. In some embodiments, the polynucleotide is incorporated into the genome of valencene-producing cells, such as valencene-producing *E. coli* cells.

The *Sr*KO or derivative in various embodiments has valencene oxidase activity. Assays for determining and quantifying valencene oxidase activity are described herein and are known in the art. Assays include expressing the *Sr*KO (or derivative) in valencene-producing cells (e.g., *E. coli* expressing FPPS and valencene synthase), and extracting the oxidized oil from the aqueous reaction media. The profile of terpenoid product can be determined quantitatively by GC/MS. Various mutations of *Sr*KO tested for effect on valencene oxidase activity are listed in **Table 2** or **Table 6**. Thus, in various embodiments, the *Sr*KO may have at least about 1, at least about 2, at least about 3, at least about 4, at least about 5, at least about 6, at least about 7, at least about 8, at least about 9, or at least about 10 mutations selected from **Table 2** or **Table 6**. In some embodiments, the *Sr*KO derivative is a modified *Sr*KO polypeptide comprising an amino acid sequence which has up to 25 mutations compared to the wild type protein according to SEQ ID NO: 37 (or its counterpart that is modified for expression in E. coli, and comprises at least the substitutions I310V, V375I or T487N in combination with at least any one or more of V375F, V375A, V375M, M120L, M120I, M120V, F129L, F129I, L114V, L114F and V121A (numbered according to SEQ ID NO:38), and optionally comprises a leader sequence (as shown in SEQ ID NO:38) supporting functional expression in *E. coli*.

Table 2.1 Summary of some *Stevia rebaudiana* kaurene oxidase mutations tested, numbered according to wild type *Sr*KO (SEQ ID NO: 37) and 8rp-t20SrKO (SEQ ID NO:38).

No.	WT	Position SEQ ID NO: 37 / SEQ ID NO:38	Mutation
1	L	59 / 47	I
2	Y	71 / 59	H
3	M	72 / 60	K
4	T	75 / 63	A
5	A	79 / 67	E
6	K	88 / 76	R
7	T	92 / 80	C
8	M	94 / 82	V
9	V	97 / 85	L
10	V	97 / 85	I
11	S	98 / 86	N
12	Q	112 / 100	S
13	N	118 / 106	K
14	K	124 / 112	T
15	A	128 / 116	R
16	T	131 / 119	S
17	M	135 / 123	T
18	M	135 / 123	Q
19	M	135 / 123	F
20	M	135 / 123	T
21	D	139 / 127	G
22	Y	141 / 129	F
23	A	152 / 140	R
24	K	161 / 149	R
25	H	162 / 150	F
26	N	183 / 171	D
27	L	192 / 180	F
28	I	195 / 183	V
29	D	220 / 208	E
30	D	244 / 232	E
31	S	279 / 267	A
32	H	284 / 272	Q
33	S	296 / 284	C
34	I	298 / 286	L
35	Q	306 / 294	K
36	Q	311 / 299	E

37	I	322 / 310	T
38	I	322 / 310	V
39	R	383 / 371	K
40	R	383 / 371	I
41	V	387 / 375	T
42	V	387 / 375	I
43	V	387 / 375	L
44	I	390 / 378	V
45	H	394 / 382	Y
46	V	400 / 388	Q
47	V	400 / 388	M
48	H	405 / 393	D
49	L	412 / 400	I
50	V	425 / 413	D
51	V	425 / 413	K
52	F	446 / 434	L
53	G	454 / 442	A
54	S	462 / 450	A
55	L	466 / 454	M
56	G	472 / 460	A
57	M	476 / 464	L
58	M	487 / 475	G
59	T	499 / 487	N
60	P	504 / 492	K
61	I	509 / 497	L
	T	499 / 487	S
62	M	135 / 123	Q
	T	499 / 487	V
63	M	135 / 123	F
	T	499 / 487	V
64	M	135 / 123	F
	T	499 / 487	F
65	M	135 / 123	F
	T	499 / 487	M
66	M	135 / 123	F
	T	499 / 487	G

Table 2.2: The following mutants were evaluated in the VO1 background (n22-yhcB-t30-VO1, SEQ ID NO:110). Positions maintain numbering of SEQ ID NO:37.

No.	WT	Position	Mutation
1	A	2	T
2	H	46	R
3	E	52	A
4	R	76	K
5	M	94	V
6	T	131	K
7	T	131	Q
8	L	150	M
9	D	191	N
10	L	231	M
11	Q	268	T
12	E	323	L
13	K	344	D
14	R	351	Q
15	I	389	L
16	I	389	V
17	I	389	A
18	I	390	L
19	I	390	M
20	V	400	Q
21	I	444	A
22	T	468	I
23	T	488	D
24	E	491	K
25	I	495	V

The *Sr*KO may be expressed in a variety of host cells, either for recombinant protein production, or for sesquiterpene (e.g., valencene) oxidation. For example, the host cells include those described in US 8,512,988, which is hereby incorporated by reference in its entirety. The host cell may be a prokaryotic or eukaryotic cell. In some embodiments the cell is a bacterial cell, such as *Escherichia spp.*, *Streptomyces spp.*, *Zymonas spp.*, *Acetobacter spp.*, *Citrobacter spp.*, *Synechocystis spp.*, *Rhizobium spp.*, *Clostridium spp.*, *Cory neb acterium spp.*, *Streptococcus spp.*, *Xanthomonas spp.*, *Lactobacillus spp.*, *Lactococcus spp.*, *Bacillus spp.*, *Alcaligenes spp.*, *Pseudomonas*

spp., *Aeromonas spp.*, *Azotobacter spp.*, *Comamonas spp.*, *Mycobacterium spp.*, *Rhodococcus spp.*, *Gluconobacter spp.*, *Ralstonia spp.*, *Acidithiobacillus spp.*, *Microlunatus spp.*, *Geobacter spp.*, *Geobacillus spp.*, *Arthrobacter spp.*, *Flavobacterium spp.*, *Serratia spp.*, *Saccharopolyspora spp.*, *Thermus spp.*, *Stenotrophomonas spp.*, *Chromobacterium spp.*, *Sinorhizobium spp.*, *Saccharopolyspora spp.*, *Agrobacterium spp.*, and *Pantoea spp.* The bacterial cell can be a Gram-negative cell such as an *Escherichia coli* (*E. coli*) cell, or a Gram-positive cell such as a species of *Bacillus*. In other embodiments, the cell is a fungal cell such as a yeast cell, such as, for example, *Saccharomyces spp.*, *Schizosaccharomyces spp.*, *Pichia spp.*, *Paffia spp.*, *Kluyveromyces spp.*, *Candida spp.*, *Talaromyces spp.*, *Brettanomyces spp.*, *Pachysolen spp.*, *Debaryomyces spp.*, *Yarrowia spp.*, and industrial polyploid yeast strains. In an embodiment, the host cell is a bacterium selected from *E. coli*, *Bacillus subtillus*, or *Pseudomonas putida*. In an embodiment, the host cell is a yeast, and may be a species of *Saccharomyces*, *Pichia*, or *Yarrowia*, including *Saccharomyces cerevisiae, Pichia pastoris,* and *Yarrowia lipolytica*.

In some embodiments, the host cell produces isopentyl pyrophosphate (IPP), which acts as a substrate for the synthesis of the sesquiterpene. In some embodiments, the IPP is produced by metabolic flux (e.g., starting with a carbon source supplied to the cell) through an endogenous or heterologous methylerythritol phosphate (MEP) or mevalonic acid (MVA) pathway. In certain embodiments, the MEP or MVA pathway may be enhanced through expression of heterologous enzymes or duplication of certain enzymes in the pathway.

The MEP (2-C-methyl-D-erythritol 4-phosphate) pathway, also called the MEP/DOXP (2-C-methyl-D-erythritol 4-phosphate/l-deoxy-D-xylulose 5-phosphate) pathway or the non-mevalonate pathway or the mevalonic acid-independent pathway refers to the pathway that converts glyceraldehyde-3-phosphate and pyruvate to IPP and DMAPP. The pathway typically involves action of the following enzymes: 1-deoxy-D-xylulose-5-phosphate synthase (Dxs), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (IspC), 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE), 2C-methyl-D-erythritol 2,4-

cyclodiphosphate synthase (IspF), 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (IspG), and isopentenyl diphosphate isomerase (IspH). The MEP pathway, and the genes and enzymes that make up the MEP pathway, are described in US 8,512,988, which is hereby incorporated by reference in its entirety. For example, genes that make up the MEP pathway include *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, and *ispA*. In some embodiments, the sesquiterpene is produced at least in part by metabolic flux through an MEP pathway, and wherein the host cell has at least one additional copy of a *dxs*, *ispD*, *ispF*, and/or *idi* gene (*e.g.*, *dxs* and *idi*; or *dxs*, *ispD*, *ispF*, and/or *idi*).

The MVA pathway refers to the biosynthetic pathway that converts acetyl-CoA to IPP. The mevalonate pathway typically comprises enzymes that catalyze the following steps: (a) condensing two molecules of acetyl-CoA to acetoacetyl-CoA (*e.g.*, by action of acetoacetyl-CoA thiolase); (b) condensing acetoacetyl-CoA with acetyl-CoA to form hydroxymethylglutaryl-CoenzymeA (HMG-CoA) (*e.g.*, by action of HMG-CoA synthase (HMGS)); (c) converting HMG-CoA to mevalonate (*e.g.*, by action of HMG-CoA reductase (HMGR)); (d) phosphorylating mevalonate to mevalonate 5-phosphate (*e.g.*, by action of mevalonate kinase (MK)); (e) converting mevalonate 5-phosphate to mevalonate 5-pyrophosphate (*e.g.*, by action of phosphomevalonate kinase (PMK)); and (f) converting mevalonate 5-pyrophosphate to isopentenyl pyrophosphate (*e.g.*, by action of mevalonate pyrophosphate decarboxylase (MPD)). The MVA pathway, and the genes and enzymes that make up the MEP pathway, are described in US 7,667,017, which is hereby incorporated by reference in its entirety.

In some embodiments, the host cell expresses a farnesyl pyrophosphate synthase (FPPS), which produces farnesyl pyrophosphate from IPP or DMAPP. As shown in **Figure 1**, farnesyl pyrophosphate is an intermediate for production of valencene. An exemplary farnesyl pyrophosphate synthase is ERG20 of *Saccharomyces cerevisiae* (NCBI accession P08524) and *E. coli* ispA. Various other prokaryotic, yeast, plant, and mammalian FPPS enzymes are known, and may be used in accordance with this aspect.

The host cell may further express a heterologous sesquiterpene synthase to produce the desired sesquiterpene, such as a valencene synthase. Several valencene synthase enzymes are known including valencene synthase from *Citrus x paradisi* or from *Citrus sinensis*. *Citrus sinensis* VS (e.g., AAQ04608.1) as well as various derivatives thereof are described in US 2012/0246767, which is hereby incorporated by reference. For example, the invention may employ an amino acid sequence of *Citrus sinensis* valencene synthase (SEQ ID NO: 12), or a derivative having from 1 to 30 mutations or from 1 to 20 or from 1 to 10 mutations with respect to the wild type amino acid sequence (SEQ ID NO:12). Such sequences may have at least 60% sequence identity, at least 70% sequence identity, at least 80% sequence identity, at least 90% sequence identity, at least 95% sequence identity, or at least about 96%, about 97%, about 98%, or about 99% sequence identity with the wild type sequence (SEQ ID NO:12). Further, valencene synthase from *Vitis vinifera* (VvVS) (SEQ ID NO: 1) has been described by Licker *et al.* (Phytochemistry (2004) 65: 2649-2659). In an embodiment, a valencene synthase comprising the amino acid sequence of VvVS or an engineered derivative thereof may be employed with the present invention. Various sesquiterpene synthase enzymes such as valencene synthase are known and are described in, for example, US 2012/0107893, US 2012/0246767, and US 7,273,735, which are hereby incorporated by reference in their entireties.

For example, in some embodiments, the valencene synthase is a VvVS derivative that comprises an amino acid sequence having from about 1 to about 40 mutations, from about 1 to about 35 mutations, from about 1 to about 30 mutations, about 1 to about 25 mutations, from about 1 to about 20 mutations, about 1 to about 15 mutations, or from about 1 to about 10 mutations independently selected from substitutions, deletions, or insertions with respect to VvVS (SEQ ID NO: 1). For example, the VvVS derivative may comprise an amino acid sequence having at least about 5 or at least about 10, but less than about 30 or about 20 mutations with respect to SEQ ID NO: 1. In various embodiments, the VvVS derivative comprises an amino acid sequence that has about 1 mutation, about 2 mutations, about 3 mutations, about 4 mutations, about 5 mutations, about 6 mutations, about 7 mutations, about 8 mutations, about 9 mutations, about 10 mutations, about 11 mutations, about 12 mutations, about

13 mutations, about 14 mutations, about 15 mutations, about 16 mutations, about 17 mutations, about 18 mutations, about 19 mutations, about 20 mutations, about 21 mutations, about 22 mutations, about 23 mutations, about 24 mutations, about 25 mutations, about 26 mutations, about 27 mutations, about 28 mutations, about 29 mutations, about 30 mutations, about 31 mutations, about 32 mutations, about 33 mutations, about 34 mutations, about 35 mutations, about 36 mutations, about 37 mutations, about 38 mutations, about 39 mutations, or about 40 mutations relative to VvVS (SEQ ID NO: 1). Such sequences may have at least 60% sequence identity, at least 70% sequence identity, at least 80% sequence identity, at least 90% sequence identity, at least 95% sequence identity, or at least about 96%, about 97%, about 98%, or about 99% sequence identity with SEQ ID NO:1. Exemplary mutations of VvVS are shown in Table 3. Mutations can be guided by a homology model of *Vitis vinifera* valencene synthase (VvVS) based on the 5-epi-aristolochene synthase crystal structure as a template (PDB: 5EAT).

Table 3.1 Summary of *Vitis vinifera* valencene synthase mutations with respect to wild type (SEQ ID NO: 1)

No.	WT	Position	Mutation
1	N	23	D
2	T	37	R
3	P	38	S
4	V	42	R
5	A	45	E
6	C	46	K
7	Q	50	R
8	K	56	E
9	K	59	R
10	R	60	K
11	K	61	M
12	T	63	R
13	T	63	K
14	N	69	Q
15	N	69	K
16	S	71	I
17	Q	72	R
18	L	73	K
19	N	75	E

20	F	76	M
21	F	76	L
22	V	80	M
23	V	80	L
24	V	85	I
25	A	86	S
26	Q	91	D
27	A	96	I
28	Q	98	E
29	Q	98	D
30	C	101	Y
31	N	102	H
32	S	103	D
33	F	104	D
34	F	104	N
35	D	111	E
36	N	116	T
37	I	117	S
38	I	117	V
39	I	117	T
40	G	120	R
41	Q	127	H
42	T	130	N
43	I	131	V
44	I	135	V
45	T	140	K
46	E	142	K
47	E	148	D
48	A	149	S
49	A	149	D
50	I	151	S
51	R	155	K
52	M	157	L
53	G	159	S
54	G	159	N
55	E	162	Q
56	E	162	K
57	A	164	S
58	V	168	T
59	G	170	D
60	L	174	M
61	A	175	E

62	A	175	D
63	K	176	E
64	T	183	K
65	A	187	S
66	M	188	L
67	E	190	N
68	G	193	K
69	A	201	S
70	N	205	E
71	R	206	Q
72	I	208	L
73	R	209	H
74	G	211	R
75	L	212	M
76	E	213	P
77	I	221	L
78	V	223	R
79	Q	225	D
80	Q	225	E
81	D	226	K
82	D	226	E
83	A	228	E
84	F	229	I
85	H	230	V
86	D	231	N
87	K	232	E
88	T	233	A
89	T	233	V
90	S	247	D
91	L	248	M
92	L	248	K
93	K	250	Q
94	E	251	K
95	S	254	K
96	N	255	E
97	A	257	S
98	K	261	A
99	E	262	D
100	D	264	G
101	Y	280	F
102	M	283	I
103	H	284	M

104	H	284	A
105	G	285	A
106	Y	287	F
107	Q	291	N
108	R	294	L
109	R	297	I
110	L	299	M
111	M	305	A
112	M	305	L
113	I	308	M
114	T	318	S
115	P	319	L
116	P	319	I
117	K	323	Q
118	R	331	K
119	D	333	E
120	I	334	E
121	I	334	V
122	N	335	K
123	N	335	Q
124	N	335	S
125	S	336	A
126	Y	343	W
127	Y	348	F
128	V	349	L
129	L	352	I
130	D	353	E
131	D	353	N
132	V	354	T
133	Y	355	F
134	K	356	E
135	K	356	N
136	I	358	V
137	E	359	D
138	E	360	Y
139	E	363	K
140	E	363	L
141	G	366	A
142	Y	369	N
143	R	370	V
144	V	371	I
145	H	372	E

146	H	372	P
147	A	374	G
148	A	374	L
149	E	376	D
150	M	378	I
151	N	380	I
152	N	380	K
153	R	383	Q
154	E	394	Q
155	E	394	D
156	E	395	N
157	E	395	G
158	H	396	Y
159	H	396	Q
160	E	402	D
161	R	405	E
162	C	414	R
163	L	415	M
164	A	417	L
165	T	418	V
166	T	419	H
167	V	422	L
168	M	424	V
169	A	428	V
170	T	429	S
171	T	437	F
172	S	438	G
173	D	439	Y
174	K	441	R
175	I	442	M
176	I	442	L
177	M	443	V
178	S	444	R
179	N	447	S
180	F	448	T
181	M	453	A
182	G	466	E
183	T	469	A
184	Q	478	E
185	Y	479	F
186	G	480	A
187	V	481	A

188	S	482	T
189	Y	487	C
190	S	488	E
191	E	489	H
192	F	490	I
193	F	490	L
194	Q	491	K
195	Q	491	N
196	Q	493	L
197	I	494	M
198	N	496	D
199	D	500	E
200	L	506	M
201	T	509	S
202	V	511	M
203	S	512	P
204	S	512	T
205	M	513	K
206	P	514	D
207	L	519	A
208	D	527	E
209	V	528	F
210	E	532	D
211	Q	533	E
212	Q	533	G
213	S	535	G
214	V	539	S
215	V	542	L
216	V	542	T
217	M	543	I
218	N	546	H
219	V	550	L
220	F	551	L
221	I	552	V
222	N	553	D
223	N	553	E
224	A	554	P
225	V	555	I

Table 3.2: Summary of mutations evaluated in the Vv2M5 background (SEQ ID NO:9).

No.	WT	Position	Mutation
226	N	18	V
227	V	21	S
228	N	23	D
229	N	27	S
230	Q	32	H
231	I	34	L
232	T	35	S
233	T	37	S
234	K	41	S
235	V	42	E
236	A	45	E
237	K	46	C
238	K	47	M
239	Q	50	R
240	I	51	V
241	D	53	E
242	K	56	E
243	V	67	A
244	A	68	N
245	N	69	D
246	N	69	Q
247	S	71	L
248	Q	72	R
249	V	80	I
250	A	86	S
251	Q	91	K
252	C	101	Y
253	N	102	D
254	N	102	H
255	M	110	D
256	D	111	E
257	G	112	D
258	I	117	S
259	T	130	N
260	R	143	E
261	R	145	N

262	A	149	S
263	S	152	N
264	G	159	N
265	G	159	S
266	V	168	T
267	K	176	E
268	K	186	E
269	A	187	S
270	S	191	H
271	Y	194	P
272	H	195	P
273	N	205	E
274	L	212	M
275	E	213	P
276	W	219	H
277	V	223	I
278	D	226	E
279	A	228	E
280	F	229	S
281	T	233	V
282	V	245	L
283	L	248	M
284	L	256	I
285	K	261	A
286	E	262	D
287	C	347	F
288	E	363	A
289	H	372	E
290	V	377	A
291	E	395	G
292	E	395	N
293	H	396	Y
294	A	399	T
295	C	414	R
296	E	426	D
297	S	438	G
298	M	443	I
299	T	469	A
300	S	488	E

301	K	491	R
302	M	513	K
303	A	517	E
304	L	519	V
305	E	532	D
306	V	550	L
307	N	553	D

Table 3.3: Summary of mutations evaluated in VS2 background (SEQ ID NO:11).

No.	WT	Position	Mutation
308	P	20	R
309	N	23	D
310	I	28	F
311	K	41	P
312	K	41	S
313	V	42	D
314	R	44	H
315	Q	50	D
316	Q	50	R
317	E	52	R
318	K	61	M
319	N	69	Q
320	Q	72	R
321	L	73	K
322	A	79	I
323	A	86	S
324	H	88	L
325	Q	91	H
326	A	96	I
327	Q	98	R
328	C	101	Y
329	N	102	H
330	C	107	F
331	I	117	S
332	G	120	L
333	T	140	K
334	R	145	N
335	S	152	V

336	V	154	I
337	V	154	P
338	R	155	K
339	M	157	L
340	G	159	M
341	A	175	D
342	K	176	E
343	A	177	P
344	L	178	I
345	H	184	Y
346	H	184	Q
347	A	187	S
348	S	191	H
349	H	195	N
350	L	196	P
351	A	201	R
352	L	212	M
353	E	213	P
354	A	217	Q
355	A	228	E
356	D	231	N
357	K	232	P
358	T	233	V
359	E	236	D
360	D	241	E
361	N	255	D
362	A	257	M
363	L	276	P
364	Y	280	F
365	M	283	I
366	V	286	A
367	T	300	M
368	T	300	I
369	T	306	L
370	T	306	I
371	L	309	I
372	A	315	V
373	E	320	D
374	K	323	R

375	S	336	T
376	E	342	D
377	C	347	L
378	A	350	I
379	N	356	H
380	E	363	G
381	Q	368	P
382	N	380	D
383	Q	381	L
384	E	395	G
385	A	407	G
386	A	407	S
387	C	414	P
388	A	417	I
389	A	432	I
390	V	436	L
391	I	442	P
392	I	442	L
393	S	445	R
394	S	446	M
395	T	450	C
396	S	458	T
397	H	459	Y
398	H	459	M
399	H	467	Q
400	T	469	A
401	E	484	P
402	Q	485	H
403	Q	485	E
404	V	486	A
405	Y	487	L
406	S	488	E
407	I	494	V
408	N	496	D
409	N	496	K
410	M	513	T
411	T	523	I
412	D	527	L
413	I	529	L

414	I	529	M
415	E	532	H
416	E	532	Y
417	S	535	A
418	R	544	H
419	N	546	Y
420	N	546	F
421	A	548	I
422	V	550	L
423	V	555	I

Thus, in various embodiments, the engineered VvVS may have at least about 1 mutation, about 2 mutations, about 3 mutations, about 4 mutations, about 5 mutations, about 6 mutations, about 7 mutations, about 8 mutations, about 9 mutations, about 10 mutations, about 11 mutations, about 12 mutations, about 13 mutations, about 14 mutations, about 15 mutations, about 16 mutations, about 17 mutations, about 18 mutations, about 19 mutations, about 20 mutations, about 21 mutations, about 22 mutations, about 23 mutations, about 24 mutations, about 25 mutations, about 26 mutations, about 27 mutations, about 28 mutations, about 29 mutations, about 30 mutations, about 31 mutations, about 32 mutations, about 33 mutations, about 34 mutations, about 35 mutations, about 36 mutations, about 37 mutations, about 38 mutations, about 39 mutations, or about 40 mutations selected from Table 3. Exemplary recombinant valencene synthases Vv1M1 (SEQ ID NO:3), Vv2M1 (SEQ ID NO:5), Vv1M5 (SEQID NO:7), Vv2M5 (SEQ ID NO:9), and VS2 (SEQ ID NO: 11) are further depicted in **Figure 3**, including an alignment in **Figure 3B**.

In certain aspects, the invention provides polynucleotides comprising a nucleotide sequence encoding a valencene synthase modified for increased expression of valencene as described above. Such polynucleotides may be expressed in host cells, either on extrachromosomal elements such as plasmids, or may be chromosomally integrated.

In various embodiments, the *Sr*KO is expressed alongside a P450 reductase to regenerate the enzyme, or alternatively, the *Sr*KO or derivative is expressed with the P450 reductase as a chimeric P450 enzyme. Functional expression of cytochrome P450

has been considered challenging due to the inherent limitations of bacterial platforms, such as the absence of electron transfer machinery and cytochrome P450 reductases, and translational incompatibility of the membrane signal modules of P450 enzymes due to the lack of an endoplasmic reticulum.

Accordingly, in some embodiments the *Sr*KO is expressed as a fusion protein with a cytochrome P450 reductase partner. Cytochrome P450 reductase is a membrane protein found in the endoplasmic reticulum. It catalyzes pyridine nucleotide dehydration and electron transfer to membrane bound cytochrome P450s. Isozymes of similar structure are found in humans, plants, other mammals, and insects. Exemplary P450 reductase partners include, for example, *Stevia rebaudiana* (*Sr*)CPR (SEQ ID NOS: 62 and 63), *Stevia rebaudiana* (*Sr*)CPR1 (SEQ ID NOS: 76 and 77), *Arabidopsis thaliana* (*At*)CPR (SEQ ID NOS: 64 and 65), *Taxus cuspidata* (*Tc*) CPR (SEQ ID NOS: 66 and 67), *Artemisia annua* (*Aa*)CPR (SEQ ID NOS: 68 and 69), *Arabidopsis thaliana* (*At*)CPR1 (SEQ ID NOS: 70 and 71), *Arabidopsis thaliana* (*At*)CPR2 (SEQ ID NOS: 72 and 73), *Arabidopsis thaliana* (*At*)R2 (SEQ ID NOS: 74 and 75); *Stevia rebaudiana* (*Sr*)CPR2 (SEQ ID NOS: 78 and 79); *Stevia rebaudiana* (*Sr*)CPR3 (SEQ ID NOS: 80 and 81); *Pelargonium graveolens* (*Pg*)CPR (SEQ ID NO: 82 and 83). Any of these P450s can be derivatized in some embodiments, for example, to introduce from 1 to about 20 mutations, or from about 1 to about 10 mutations. **Figure 6B** shows an alignment of amino acid sequences for *Arabidopsis thaliana* and *Artemisia annua* CPR sequences (SEQ ID NOS:72, 74, 68, 64, and 70). **Figure 6C** shows an alignment of *Stevia rebaudiana* CPR sequences (SEQ ID NOS: 78, 80, 62, and 76). Figure 6D shows an alignment of eight CPR amino acid sequences (SEQ ID NO: 74, 72, 82, 68, 80, 62, 78, and 76).

Engineering of P450 fusion proteins is disclosed, for example, in US 2012/0107893 and US 2012/0164678, both of which are hereby incorporated by reference in their entireties. In certain embodiments, the *Sr*KO is fused to the cytochrome P450 reductase partner through a linker. Exemplary linker sequences, which are predominantly serine, glycine, and/or alanine, and optionally from one to five charged amino acids such as lysine or arginine, include, for example, GSG,

GSGGGGS (SEQ ID NO: 113), GSGEAAAK (SEQ ID NO: 114), GSGEAAAKEAAAK (SEQ ID NO: 115), GSGMGSSSN (SEQ ID NO: 116), and GSTGS (SEQ ID NO: 117). The linker is generally flexible, and contains no more than one, two, or three hydrophobic residues, and is generally from three to fifty amino acids in length, such as from three to twenty amino acids in length. In other embodiments, a P450 reductase is expressed in the host cell separately, and may be expressed in the same operon as the *Sr*KO in some embodiments. In some embodiments, the P450 reductase enzyme is expressed separately in the host cell, and the gene is optionally integrated into the genome or expressed from a plasmid.

In certain embodiments the N-terminus of the P450 enzymes may be engineered to increase their functional expression. The N-terminus of membrane-bound P450 plays important roles in enzyme expression, membrane association and substrate access. It has been reported that the use of rare codons in the N-terminus of P450 significantly improved the expression level of P450. Further, since most plant P450 enzymes are membrane-bound and hydrophobic substrates are thought to enter the enzymes through channels dynamically established between the P450 and membrane, N-terminal engineering can affect the association of the membrane and P450 and therefore the access of substrate to the enzyme. Accordingly, in an embodiment, N-terminal engineering of *Sr*KO generates an *Sr*KO derivative that either maintains or shows enhanced valencene oxidase activity in a host system such as *E. coli* or yeast. An exemplary N-terminal sequence is MALLLAVF (SEQ ID NO:112), and other exemplary sequences include sequences of from four to twenty amino acids (such as from four to fifteen amino acids, or from four to ten amino acids, or about eight amino acids) that are predominately hydrophobic, for example, constructed predominately of (at least 50%, or at least 75%) amino acids selected from leucine, valine, alanine, isoleucine, and phenylalanine.

In some embodiments, the *Sr*KO is a derivative having a deletion of at least a portion of its N-terminal transmembrane region, and the addition of an inner membrane transmembrane domain from *E. coli* yhcB or derivative thereof. In these embodiments, the P450 enzyme has a more stable and/or productive association with the *E. coli* inner

membrane, which reduces cell stress otherwise induced by the expression of a membrane-associated P450 enzyme. In some embodiments, the *Sr*KO is a derivative having a deletion of from 15 to 35 amino acids of its N-terminal transmembrane domain, and the addition of from 15 to 25 amino acids of the transmembrane domain from *E. coli* yhcB or derivative thereof. In some embodiments, the N-terminal transmembrane domain of the derivative comprises the amino acid sequence MAWEYALIGLVVGIIIGAVA (SEQ ID_NO:118), or an amino acid sequence having from 1 to 10 or from 1 to 5 amino acid mutations with respect to SEQ ID NO:118.

In some embodiments, the host cell further expresses one or more enzymes that divert product toward nootkatone. For example, the host cell may express an alcohol dehydrogenase enzyme producing nootkatone from nootkatol, examples of which include *Rhodococcus erythropolis* CDH (SEQ ID NO: 84), *Citrus sinensus* DH (SEQ ID NO: 86), *Citrus sinensus* DH1 (SEQ ID NO: 88), *Citrus sinensus* DH2 (SEQ ID NO: 90), *Citrus sinensus* DH3 (SEQ ID NO: 92), *Vitis vinifera* DH (SEQ ID NO: 94), *Vitis vinifera* DH1 (SEQ ID NO: 96), *Citrus sinensus* ABA2 (SEQ ID NO: 98), *Brachypodium distachyon* DH (SEQ ID NO: 100), and *Zingiber zerumbet* SDR (SEQ ID NO: 102). The alcohol dehydrogenase may comprise an amino acid sequence having at least 70%, at least 80%, or at least 90% sequence identity to one or more of the enzymes described in this paragraph, and with the activity of converting nootkatol to nootkatone.

Sesquiterpenes (e.g., valencene and its oxygenated products) can be produced as biosynthetic products of the non-mevalonate pathway in *E. coli* comprising two modules: the native upstream pathway forming Isopentenyl Pyrophosphate (IPP) and a heterologous downstream terpenoid-forming pathway. A multivariate-modular approach to metabolic pathway engineering can be employed to optimize the production of sesquiterpenes in an engineered *E. coli*. The multivariate-modular pathway engineering approach is based on a systematic multivariate search to identify conditions that optimally balance the two pathway modules to minimize accumulation of inhibitory intermediates and flux diversion to side products.

WO 2011/060057, US 2011/0189717, US 2012/107893, and US 8512988 (each of which are hereby incorporated by reference) describe methods and compositions for optimizing production of terpenoids in cells by controlling expression of genes or proteins participating in an upstream pathway and a downstream pathway. This can be achieved by grouping the enzyme pathways into two modules: an upstream (MEP) pathway module (e.g., containing one or more genes of the MEP pathway) and a downstream, heterologous pathway to sesquiterpene production. Using this basic configuration, parameters such as the effect of plasmid copy number on cell physiology, gene order and promoter strength in an expression cassette, and chromosomal integration are evaluated with respect to their effect on terpene and terpenoid (e.g., sesquiterpene) production. Expression of genes within the MEP pathway can thus be regulated in a modular method. As used herein, regulation by a modular method refers to regulation of multiple genes together. By way of example, multiple genes within the MEP pathway can be recombinantly expressed on a contiguous region of DNA, such as an operon. It should be appreciated that modules of genes within the MEP pathway, consistent with aspects of the invention, can contain any of the genes within the MEP pathway, in any order. In some embodiments, a gene within the MEP pathway is one of the following: dxs, ispC, ispD, ispE, ispF, ispG, ispH, idi, ispA or ispB. A non-limiting example of a module of genes within the MEP pathway is a module containing the genes dxs, idi, ispD and ispF, and referred to as dxs-idi-ispDF.

The manipulation of the expression of genes and/or proteins, including modules such as the dxs-idi-ispDF operon, and a FPPS-VS operon, can be achieved through methods known to one of ordinary skill in the art. For example, expression of the genes or operons can be regulated through selection of promoters, such as inducible promoters, with different strengths. Several non-limiting examples of promoters include Trc, T5 and T7. Additionally, expression of genes or operons can be regulated through manipulation of the copy number of the gene or operon in the cell.

The expression of one or more genes and/or proteins within the MEP pathway can be upregulated and/or downregulated. In certain embodiments, upregulation of one

or more genes and/or proteins within the MEP pathway can be combined with downregulation of one or more genes and/or proteins within the MEP pathway. By way of example, in some embodiments, a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway is used, at least in part, to amplify isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), substrates of GGPPS. In some embodiments, overexpression of one or more components of the non-mevalonate (MEP) pathway is achieved by increasing the copy number of one or more components of the non-mevalonate (MEP) pathway. In this regards, copy numbers of components at rate-limiting steps in the MEP pathway such as (*dxs*, *ispD*, *ispF*, *idi*) can be amplified, such as by additional episomal expression.

In some embodiments, the production of indole is used as a surrogate marker for sesquiterpene production, and/or the accumulation of indole in the culture is controlled to increase sesquiterpene production. For example, in various embodiments, accumulation of indole in the culture is controlled to below about 100 mg/L, or below about 75 mg/L, or below about 50 mg/L, or below about 25 mg/L, or below about 10 mg/L. The accumulation of indole can be controlled by balancing protein expression and activity using the multivariate modular approach described above, and/or is controlled by chemical means.

In other aspects, the invention provides a method for making a product containing an oxygenated sesquiterpene (as described), which comprises incorporating the oxygenated sesquiterpene prepared and recovered according to the method described above into a consumer or industrial product. For example, the product may be a flavor product, a fragrance product, a cosmetic, a cleaning product, a detergent or soap, or a pest control product (e.g., an insect repellant). In some embodiments, the oxygenated product recovered and optionally enriched by fractionation (e.g. fractional distillation) is nootkatone, and the product is a flavor product selected from a beverage, a chewing gum, a candy, or a flavor additive, or is an insect repellant.

The oxidized product can be recovered by any suitable process, including partitioning the desired product into an organic phase. The production of the desired product can be determined and/or quantified, for example, by gas chromatography (e.g.,

GC-MS). The desired product can be produced in batch or continuous bioreactor systems. Production of product, recovery, and/or analysis of the product can be done as described in US 2012/0246767, which is hereby incorporated by reference in its entirety. For example, in some embodiments, oxidized oil is extracted from aqueous reaction medium, which may be done by partitioning into an organic phase, e.g., using an organic solvent such as an alkane such as heptane, followed by fractional distillation. Sesquiterpene and sesquiterpenoid components of fractions may be measured quantitatively by GC/MS, followed by blending of the fractions to generate a desired nootkatone-containing ingredient for flavour (or other) applications.

In other aspects, the invention provides polynucleotides comprising a nucleotide sequence encoding a P450 derivative described herein. The polynucleotide may be codon optimized for expression in *E. coli* or yeast in some embodiments. In another example, the polynucleotide may comprise a nucleotide sequence encoding a *Sr*KO fusion protein, optionally with a P450 reductase partner as described herein. In other embodiments, the invention provides polynucleotides comprising a nucleotide sequence encoding a sesquiterpene synthase variant described herein, which may likewise be codon optimized for expression in *E. coli* or yeast. Such polynucleotides may further comprise, in addition to sequences encoding the P450 or sesquiterpene synthase, one or more expression control elements. For example, the polynucleotide may comprise one or more promoters or transcriptional enhancers, ribosomal binding sites, transcription termination signals, and polyadenylation signals, as expression control elements. The polynucleotide may be inserted within any suitable vector, including an expression vector, and which may be contained within any suitable host cell for expression. The polynucleotide may be designed for introduction and/or protein expression in any suitable host cell, including bacterial cells and yeast cells, and may be expressed from a plasmid, or may be chromosomally integrated. In some embodiments, the recombinant nucleic acid molecule encodes an *Sr*KO derivative with a higher activity for oxidation of valencene than the wild type enzyme (SEQ ID NO:37), and having a leader sequence as described, such as the leader sequence MALLLAVF (SEQ ID NO:117) or leader sequence derived from *E. coli* yhcB. In certain embodiments, the recombinant nucleic acid molecules further encodes either as an operon or as a fusion in frame with the

*Sr*KO derivative, an *Sr*CPR or derivative thereof capable of regenerating the *Sr*KO enzyme. When present as a fusion protein, the SrKO derivative and the *Sr*CPR may be connected by a linking sequence of from 3 to 10 amino acids (e.g., 5 amino acids). In some embodiments, the linking sequence is predominately glycine, serine, and/or alanine and may comprise the sequence GSTGS.

In other aspects, the invention provides host cells producing an oxygenated sesquiterpene as described herein, and which express all of the enzyme components for producing the desired oxygenated sesquiterpene from isopentyl pyrophosphate (IPP). For example, the host cell in various embodiments expresses a farnesyl pyrophosphate synthase, a sesquiterpene synthase, and the *Sr*KO or derivative thereof. IPP may be produced through the MEP and/or MVA pathway, which may be endogenous to the host cell or modified through expression of heterologous enzymes or duplication of certain enzymes in the pathway. Host cells include various bacteria and yeast as described herein.

In still other aspects, the invention provides sesquiterpene products produced by the methods and host cells described herein. As disclosed herein, *Sr*KO enzyme showed unique activities by creating nootkatol and further oxidizing to the ketone, nootkatone, and produced different oxygenated terpene products including hydroxygermacra-1(10)5-diene, and murolan-3,9(11) diene-10-peroxy.

Further, other P450 enzymes tested, including previously known sesquiterpene CYP450's or P450's having hydroxylating activity on the valencene substrate produced one of the stereoisomers (beta nootkatol) and only minor amounts of the ketone (nootkatone). Specifically, the other sesquiterpene CYP450 enzymes produced beta-nootkatol and hydroxyl valencene as major products, while Taxol CYP450 enzyme did not produce any oxygenated valencene (**Table 4** and **Figure 7**).

In certain aspects, the invention relates to *Sr*KO derivative enzymes. For example, the *Sr*KO derivative may comprise an amino acid sequence that has one or more mutations at positions selected from 46, 76, 94, 131, 231, 284, 383, 390, 400, 444, 468, 488 and 499 relative to SEQ ID NO:37. For example, in some embodiments

the *Sr*KO is a derivative comprising an amino acid sequence having one or more (two, three, four, or all) of the mutations selected from R76K, M94V, T131Q, F231L, H284Q, R383K, I390L, T468I, and T499N relative to SEQ ID NO:37. In some embodiments, the *Sr*KO derivative comprises an amino acid sequence selected from SEQ ID NOS:55-61, 104, or 105 which were engineered according to this disclosure to improve activity for oxygenation of valencene (e.g., production of nootkatone). In some embodiments, the derivative comprises an amino acid sequence having from one to twenty mutations, or one to ten mutations, or one to five mutations relative to a sequence selected from SEQ ID NOS: 55-61, 104, or 105 with the proviso that the amino acid sequence has one, two, three or more mutations at positions selected from 46, 76, 94, 131, 231, 284, 383, 390, 400, 444, 468, 488 and 499 relative to SEQ ID NO:37, or the proviso that the *Sr*KO derivative comprises an amino acid sequence having one, two, three or more (or all) of the mutations selected from R76K, M94V, T131Q, F231L, H284Q, R383K, I390L, T468I, and T499N relative to SEQ ID NO:37. As shown herein, these mutations increase the level of *Sr*KOs valencene oxidation activity.

In these or other embodiments, the *Sr*KO is a derivative having a deletion of at least a portion of its N-terminal transmembrane region, and the addition of an inner membrane transmembrane domain from *E. coli* yhcB or derivative thereof. In some embodiments, the *Sr*KO is a derivative having a deletion of from 15 to 35 amino acids of its N-terminal transmembrane domain (relative to SEQ ID NO:37), and the addition of from 15 to 25 amino acids of the transmembrane domain from *E. coli* yhcB or derivative thereof. In some embodiments, the N-terminal transmembrane domain of the derivative comprises the amino acid sequence MAWEYALIGLVVGIIIGAVA (SEQ ID NO:118), or an amino acid sequence having from 1 to 10 or from 1 to 5 amino acid mutations with respect to SEQ ID NO:118.

In still other aspects, the invention provides a method of preparing the modified *Sr*KO polypeptide, wherein the method comprises the steps of: (i) culturing a host cell expressing the modified polypeptide under conditions which permit expression of the polypeptide; and (ii) optionally recovering the polypeptide.

In still other aspects, the invention provides a method of producing an oxygenated sesquiterpene comprising the steps of: (i) providing the modified *Sr*KO polypeptide, (ii) contacting a sesquiterpene with the modified *Sr*KO polypeptide, and (iii) recovering the produced oxygenated sesquiterpene. The method may further comprise providing a CPR enzyme for regenerating the *Sr*KO cofactor (e.g., *Sr*CPR). In some embodiments, the oxygenated sesquiterpene is recovered as an oil. In some embodiments, the sesquiterpene is valencene. In some embodiments, the oxygenated sesquiterpene comprises hydroxygermacra-1(10)5-diene, murolan-3,9(11) diene-10-peroxy, nootkatol, and nootkatone. In some embodiments, the predominant oxygenated product is nootkatone and/or nootkatol.

In another aspect, there is provided an *Sr*KO crystal model structure (CMS) based on the structural coordinates of P45017A1, with an amino acid sequence of *Sr*KO or derivative described herein. The CMS comprises a terpene binding pocket domain (TBD) that comprises a terpene binding pocket (TBP) and a terpene (e.g., valencene) bound to the TBD. **Figures 8A** and **8B**. This *Sr*KO crystal model structure (CMS) facilitates in-silico testing of *Sr*KO derivatives.

Thus, in still other embodiments, the invention provides a method of screening for a terpene capable of binding to a TBD wherein the method comprises the use of the *Sr*KO CMS. In another aspect, the invention provides a method for screening for a terpene capable of binding to the TBP, and the method comprises contacting the TBP with a test compound, and determining if said test compound binds to said TBP. In some embodiments, the method is to screen for a test compound (e.g., terpenes) useful in modulating the activity of a *Sr*KO enzyme.

In another aspect, the invention provides a method for predicting, simulating or modelling the molecular characteristics and/or molecular interactions of a terpene binding domain (TBD) comprising the use of a computer model, said computer model comprising, using or depicting the structural coordinates of a terpene binding domain as defined above to provide an image of said ligand binding domain and to optionally display said image.

EXAMPLES

Example 1: Construction of sesquiterpene precursor (Valencene) producing E. coli strain

E. coli overexpressing upstream MEP pathway genes *dxs*, *ispD, ispF*, and *idi* was created, which facilitates flux to the isoprenoid precursor isopentyl-pyrophosphate (IPP) supporting more than 1 g/L titers of a heterologous diterpenoid product (3). Strains were constructed producing a variety of terpenoids including mono- and sesquiterpenes by replacing the geranylgeranyl pyrophosphate synthase (GGPS) and diterpene synthase with a farnesyl pyrophosphate synthase (FPPS) and sesquiterpene synthase or a geranyl pyrophosphate synthase (GPPS) and monoterpene synthase. For developing a sesquiterpene producing strain to test the CYP450s for novel oxygenated terpenes, a valencene synthase enzyme was cloned and expressed in the MEP pathway overexpressed *E. coli* strain. The high substrate flux helps identify the activity of the CYP450. Previously, research on an oxygenated taxadiene producing strain showed a significant drop in the productivity upon transferring the CYP450 pathway to the taxadiene producing strain (300 mg/L to ~10 mg/L).

Further, multivariate modular metabolic engineering (MMME) was applied for balancing the pathway for high level production of valencene. Naturally occurring valencene synthases, such as that from *Vitis vinifera*, often perform sub-optimally (~5mg/L) even after MMME optimization, compared to previous results obtaining 100's of mg/L diterpenoids. Enzymes involved in the sesquiterpene biosynthesis can be difficult to express in *E. coli*, and also are deficient in kinetics relative to those involved in primary metabolism (17).

A homology model for the *Vitis vinifera* valencene synthase (VvVS) was constructed using the BioLuminate® software package (Schrodinger, Inc.) with the 5-epi-aristolochene synthase crystal structure as a template (PDB: 5EAT). Further, to identify the natural mutational landscape of terpene synthases, an extensive multiple sequence alignment incorporating hundreds of related terpene synthase sequences was created. Using this information, mutations were designed using a combination of back-to-consensus, *in silico* energetics, and structural analysis. Back-to-consensus mutations have been shown to be an important tool for improving stability (19,20) and expression

(21). Energetics calculations based on atomic force-field models in BioLuminate were used to assess the ΔΔG of folding for individual mutations predicted for positions with low solvent-accessible surface area, which were predicted to affect folding and stability.

By applying the MMME approach, a balanced upstream and downstream valencene production strain was identified incorporating a codon-optimized version of VvVS on a plasmid with a p15A origin of replication and a T7 promoter. This strain background was then used to screen designed synthase enzyme mutations. Using the aforementioned protein engineering tools we designed over 200 unique point mutations (**Table 3**) which were then constructed in the p15A-T7 screening plasmid using site-directed mutagenesis. Mutated enzyme variants were transformed into the screening strain, triplicate colonies were cultured in selective LB cell culture medium overnight, and then inoculated into a minimal R-medium and cultured for four days at 22°C. Cultures were extracted using methyl tert-butyl ether (MTBE) and analyzed by combined gas chromatography/mass spectrometry for productivity of valencene.

Approximately one-fifth of the designed point mutations increased valencene productivity in our screening strain by at least 20% (**Figure 2**). Beneficial point mutations were then strategically combined to confer increasingly advantageous phenotypes. Recombined valencene synthase sequences are provided as **Vv1M1** (Mutations- R331K, I334E, N335S, V371I, A374L, T418V, S482T, S512P, K356N, Q491K, E394D, A428V, Y348F, T318S, L352I, I442L, A554P), **Vv2M1** (Mutations- R331K, I334E, N335S, V371I, A374L, T418V, S482T, S512P, K356N, Q491K, E394D, A428V, V542T, G480A, M305L, K441R, A554P), **Vv1M5** (Mutations- R331K, I334E, N335S, V371I, A374L, T418V, S482T, S512P, K356N, Q491K, E394D, A428V, Y348F, T318S, L352I, I442L, A554P, H284M, C46K, F448T, Q533E), and **Vv2M5** (Mutations- R331K, I334E, N335S, V371I, A374L, T418V, S482T, S512P, K356N, Q491K, E394D, A428V, V542T, G480A, M305L, K441R, A554P, H284M, C46K, F448T, Q533E) (**Figure 3**). When either of these enzymes was overexpressed in our MEP pathway strain with dxs-idi-ispDF overexpressed, and balanced using MMME, the titers of valencene obtained were sufficient to motivate incorporation of P450 enzymes to test their ability to catalyze the formation of

oxygenated valencene. Titers of valencene before P450 incorporation were about 30 mg/L.

Example 2: Functional activity of CYP450 library on valencene scaffold

Valencene was used as a model system to validate the power of CYP450-based oxygenation chemistry for production terpene chemicals.

The CYPP450 candidate screening was conducted using the valencene producing *E. coli* strains as host background. For constructing the CYP450 for functional expression, a proprietary plasmid system, p5Trc (plasmid derived from pSC101) was used to construct a plasmid containing the candidate P450 fused to an N-terminal truncated *Stevia rebaudiana* cytochrome P450 reductase (*Sr*CPR) through a flexible 5-amino acid linker (GSTGS, SEQ IDNO:117). The sequences of the various candidate P450s are shown in **Figure 4**. The candidate CYP450's were analyzed for N-terminal membrane associating regions which were truncated and a 8-amino acid leader sequence (MALLLAVF, SEQ ID NO:112) was added to the fusion (**Figures 5A** and **5B**). CPR red/ox partners from *Arabidopsis thaliana* and *Taxus cuspidata* were also prepared in similar genetic constructions. Since the native *Sr*CPR was effective, the level of activity of these constructs was not determined. The sequences of the various CPR red/ox partners are shown in **Figure 6**. Following transformation of p5Trc-CYP450-L-*Sr*CPR to valencene producing strain, the strains were cultured overnight at 30°C in antibiotic selective LB media. These cultures were then used to inoculate 2 mL antibiotic selective R-media cultures in hungate tubes with 15g/L glycerol and 0.1mM IPTG which were subsequently cultured for 4-days at 22°C before being extracted with methyl tert-butyl ether (MTBE).

A set of CYP450 enzymes, from those listed in **Table 4**, was selected and classified for both sesqui- and diterpene oxygenation in this *E. coli* system. Among the various CYP450 enzymes tested for oxygenation on valencene, kaurene oxidase from *Stevia rebaudiana* (*Sr*KO) (16) was discovered to have a unique oxygenation chemistry on the valencene scaffold. *Sr*KO natively oxidizes the diterpene (-)-kaurene at the C19 position to (-)-kaurenoic acid. *Sr*KO enzyme showed unique activities in the present studies by creating different stereoisomers of the hydroxylated product (alpha and beta

nootkatol and further oxidizing to the ketone, nootkatone), and produced different oxygenated terpene products including hydroxygermacra-1(10)5-diene, murolan-3,9(11) diene-10-peroxy, in addition to the alpha-nootkatol, beta-nootkatol, and nootkatone. Other P450's, including the previously known sesquiterpene CYP450's for hydroxylating valencene produced only one of the isomers (beta nootkatol) and only detectable amounts of ketone (nootkatone). Specifically, the other sesquiterpene CYP450 enzymes produced beta-nootkatol and hydroxyl valencene as major products, while another diterpene CYP450 enzymes (e.g., Taxus 5-alpha hydroxylase) produced nootkatol as only a minor (detectable) product (**Table 4** and **Figure 7**).

Table 4: Major Products Formed From Valencene by Select P450 Enzymes in E. coli

SPECIES	NAME	MAJOR PRODUCTS
Cichorium intybus	CiVO	β-nootkatol, α-cadinol, hydroxyl valencene.
Hyoscyamus muticus	HmPO	β-nootkatol, α-cadinol, hydroxyl valencene, nootkatone
Latuca spicata	LsGAO	β-nootkatol, α-cadinol, isovalencenol, nootkat-11-en-10-ol.
Barnadesia spinosa	BsGAO	β-nootkatol, α-cadinol, isovalencenol.
Nicotiana tabacum	NtEAO	α-cadinol, nootkat-11-en-10-ol.
Stevia rebaudiana	*Sr*KO	α-nootkatol, hydroxygermacra-1(10)5-diene, β-nootkatol, nootkatone, murolan-3,9(11) diene-10-peroxy
Zingiber zerumbet	ZzHO	α-cadinol, nootkat-11-en-10-ol.
Citrus x paradisi	CpVO	α-cadinol, nootkat-11-en-10-ol.
Mentha spicata	MsL6OH	α-cadinol.
Nicotiana tabacum	NtVO	α-cadinol, nootkat-11-en-10-ol.
Solanum tuberosum	StVO	α-cadinol, β-nootkatol, globulol.
Arabidopsis thaliana	AtKO	α-cadinol, nootkat-11-en-10-ol.
Cichorium intybus	Ci2VO	β-nootkatol, α-cadinol, isovalencenol.
Artemesia annua	AaAO	α-cadinol, murolol, nootkat-11-en-10-ol.

Taxus 5-alpha hydroxylase P450		α-cadinol

Example 3: Structural and mutational studies of SrKO

Once the unique activities of *Sr*KO were identified, experiments were conducted to improve its ability to conduct its diverse oxidation of valencene. The crystal structure for *Sr*KO has not been described. Blast search of *Sr*KO against RCSB Protein Data Bank shows the sequence identity of *Sr*KO to P450 enzymes with crystal structures are low (~20%). Given the conservative folding structures of P450s regardless of its low sequence identity, state-of-the-art protein modeling tools were used to build on *Sr*KO. The crystal structure of membrane-bound cytochrome P450 17 A1 (*see* DeVore N.M., Scott E.E., *Nature*, 482, 116-119, 2012) which catalyses the biosynthesis of androgens in human was selected as the template for model development. Using BioLuminate protein modeling software, a homology model was developed (**Figure 8A**) such that the positioning of key residues and characteristic motifs (*see* Gotoh O., *J. Biol Chem*, 267, 83-90, 1992) aligned well with the template. Furthermore, a homology model for *Sr*KO which included the prosthetic heme-iron complex was also constructed. AutoDock VINA was then used to create an ensemble of possible binding modes for valencene in the *Sr*KO active site (**Figure 8B**) (29).

In addition, a Blast search of *Sr*KO against NCBI non-redundant protein sequence library returned no orthologs with sequence identity greater than 80% (except the *Sr*KO itself). The top hits are listed in the **Table 5**.

Table 5: BLAST search with SrKO in preparation of Homology Model

Enzyme Name	Species	Sequence Identity	Accession
kaurene oxidase	Stevia rebaudiana	99%	AAQ63464.1
ent-kaurene oxidase 2	Lactuca sativa	79%	BAG71198.1
ent-kaurene oxidase 1	Lactuca sativa	71%	BAG71197.1
ent-kaurene oxidase	Ricinus communis	63%	XP_002510288.1

Once the unique activities of *Sr*KO were identified, experiments were conducted to improve its ability to conduct its diverse oxidation of valencene. Using the back-to-consensus mutagenesis strategy, a multiple sequence alignment of P450

enzymes was constructed including sequences (after clustering and elimination of sequences with greater than 90% identity) from a BLAST search of the Uniref100 database using 4 seed kaurene oxidase genes, from a BLAST search of the bacterial proteome using $P450_{BM3}$, $P450_{CAM}$, and $P450_{eryF}$ as seed genes, and the most closely related *Sr*KO homologs. Based on the homology model, the multiple sequence alignment, and the literature, various point mutations and double mutations were designed and tested. These cytochrome P450 derivatives were assessed for improvements in total oxygenated terpene productivity (e.g., total of the major peaks observed by GC/MS) in the *in vivo* testing system described above. Mutagenesis on active site positions guided by the model revealed several variants with significantly improved oxygenated products (**Table 6** and **Table 7** below).

Table 6: Binding pocket mutations and its fold productivity of total oxygenated oil as compared to the *Sr*KO of SEQ ID NO: 38

Mutant # (Table 2)	Mutants (numbered according to SEQ ID NO:38)	Fold productivity
38	I310V	1.5
37	I310T	0.0
42	V375I	1.4
41	V375T	0.0
19	M123F	0.0
20	M123T	0.3
18	M123Q	0.0
59	T487N	2.5
66	M123F_T487G	0.0
63	M123F_T487V	0.0
62	M123Q T487V	0.0
59	T487N_V375F	2.2
59	T487N_V375A	1.8
59	T487N_V121A	2.0
59	T487N_V375M	1.9
59	T487N_M120L	1.8
59	T487N_M120I	1.8
59	T487N_L114V	1.4
59	T487N_F219L	3.5

59	T487N_M120V	1.1
59	T487N_F219I	3.3
59	T487N_L114F	1.2

Table 7: Non-binding pocket point mutations and productivity of total oxygenated oil compared to the *Sr*KO of SEQ ID NO: 38

Mutant # (Table 2)	**Mutants (numbered according to SEQ ID NO:38)**	**Fold productivity**
53	G442A	0.849153
55	L454M	0.717318
44	I378V	0.349005
47	V388M	0.792428
9	V85I	0.795913
51	V413K	0.902039
60	P492K	0.131657
40	R371I	0.657808
7	T80C	0.342501
23	A140R	0.014872
2	Y59H	0.406787
5	A67E	0.937429
8	M82V	1.585588
11	S86N	0.977752
22	Y129F	0.686276
24	K149R	0.990776
29	D208E	0.853446
31	S267A	0.79152
32	H272Q	0.958227
33	S284C	0.652348
39	R371K	1.443497
45	H382Y	0.609951
46	V388Q	0.924043
49	L400I	0.682775
50	V413D	0.039261
52	F434L	0.793926
57	M464L	0.689696
58	M475G	0.573906
61	I497L	0.679949
15	A116R	0.216353

1	L47I	0.88992
25	H150F	0.666723

Example 4: Isolation and Evaluation of Nootkatone

The product derived from oxidation of valencene by the cytochrome P450 enzyme *Sr*KO (SEQ ID NO:38) was analysed by GC/MS (Agilent 6800; Column: Rtx-5, 0.32 mm x 60 m x 1.0 μm film thickness; GC Temp. Program: 40°C for 5 min, increased at 4°C/min to 300°C and held for 30 min.) resulting in the data provided in Table 8A and 8B.

Table 8A. SrKO oxidation of valencene

Ret.Time	**Compound Name**	**CAS #**	**GC-FID Area %**
33.762	dodecane	112-40-3	6.70
35.440	glyceryl diacetate I		5.26
38.767	triacetin	102-76-1	4.48
39.518	unknown		10.17
40.176	unknown		7.52
42.012	unknown		2.53
44.437	unknown		20.25
44.816	valencene	4630-07-3	1.97
45.546	nootkatene	5090-61-9	1.09
46.260	unknown		2.14
46.395	unknown		6.77
46.869	unknown		4.01
47.394	germacrene D-4-ol	74841-87-5	1.23
48.273	unknown		1.86
49.659	T-muurolol	19912-62-0	0.69
49.753	an unknown sesquiterpene		0.56
50.336	an unknown sesquiterpene		0.58
51.025	epinootkatol (or alpha nootkatol)	50763-66-1	1.96
51.430	Nootkatol (or beta nootkatol)	50763-67-2	3.54
54.138	nootkatone	4674-50-4	15.87
54.501	6-isopropenyl-4,8a-dimethyl-4a,5,6,7,8,8a-hexahydro-2(1H)-naphthalenone	76784-84-4	0.84
		TOTAL	**100.00**

Table 8B. *Sr*KO oxidation of valencene

Ret.Time	Compound Name	CAS #	GC-FID Area %
33.763	dodecane	112-40-3	7.26
35.470	glyceryl diacetate I		7.07
38.773	triacetin	102-76-1	6.19
39.526	unknown		11.56
40.179	unknown		8.10
44.440	unknown		23.95
44.821	valencene	4630-07-3	6.88
45.545	nootkatene	5090-61-9	2.22
46.404	unknown		5.08
46.879	unknown		3.66
47.399	germacrene D-4-ol	74841-87-5	2.89
48.279	unknown		2.27
49.665	T-muurolol	19912-62-0	0.94
50.342	an unknown sesquiterpene		1.71
51.027	epinootkatol	50763-66-1	2.48
51.444	nootkatol	50763-67-2	5.24
54.152	nootkatone	4674-50-4	2.49
		TOTAL	**100.00**

Similar analysis was conducted on the product produced by *Sr*KO derivatives. It was confirmed that product profiles are comparable, and that the major products of nootkatone, nootkatol can be produced at higher levels based on mutagenesis of *Sr*KO.

The oxidized oil product can then be extracted from the aqueous reaction medium using an appropriate solvent (e.g., heptane) followed by fractional distillation. The chemical composition of each fraction can be measured quantitatively by GC/MS. Fractions can be blended to generate the desired nootkatol and/or nootkatone ingredients for use in flavour or other applications.

Verification of acceptability can be carried out by direct comparison to a reference nootkatone flavouring product (for example, an existing natural flavouring commercial product obtained from Frutarom) with analysis provided in Table 9.

Table 9. Analysis of commercially available natural flavouring nootkatone from Frutarom

Ret.Time	Compound Name	CAS #	GC-FID Area %
42.307	limonene glycol	1946-00-5	0.201
42.792	decanoic acid	334-48-5	0.115
49.405	valencene	4630-07-3	0.039
50.362	delta-cadinene	483-76-1	0.268
52.757	alpha-elemol	639-99-6	2.178
53.11	spathulenol	6750-60-3	0.264
53.423	caryophyllene oxide	1139-30-6	0.394
53.748	viridiflorol	552-02-3	0.061
54.225	unknown sesquiterpenoid (MW = 220, tent)		0.113
54.853	unknown		2.985
55.386	unknown		2.251
55.97	T-muurolol	19912-62-0	0.399
56.192	bulnesol	22451-73-6	0.722
56.523	7(11), 4b-selinenol; tentative		1.425
56.65	unknown (MW = 232, tent)		0.663
56.937	beta-sinensal	3779-62-2	0.914
57.449	unknown		0.285
57.589	cedrenal; tentative		0.438
58.189	unknown sesquiterpenoid(s)		1.077
58.73	unknown sesquiterpenoids (MW = 220, 222, tent)		0.537
59.102	beta,gamma-nootkatone	35936-67-5	1.805
59.32	myristic acid	544-63-8	0.058
59.537	1,10-dihydronootkatone	20489-53-6	0.582
59.75	a nootkatone isomer		0.442
60.507	nootkatone isomers (2); tentative		0.812
60.782	unknowns (2)		0.605
61.034	hexadecanal	629-80-1	0.302
62.93	nootkatone	4674-50-4	74.287
63.057	3,11-eudesmadiene-2-one (5S,7R,10R)	86917-81-9	1.909
63.14	unknown (MW = 234, tent)		0.18
63.26	unknown (MW = 232, tent)		0.105
64.112	heptadecanal	629-90-3	0.344
64.403	unknown sesquiterpenoid		0.446
65.16	unknown sesquiterpenoid		0.147
65.384	palmitic acid	57-10-3	0.154
65.599	alpha-camphorene	532-87-6	0.249
65.75	unknown(s)		0.054
65.878	dehydro-alpha-vetivenone; tentative		0.115
66.056	nootkatone, 9-oxo	86925-44-2	0.172
66.371	ethyl palmitate	628-97-7	0.185

66.856	cis-9-octadecenal	2423-10-1	0.239
66.986	unknown sesquiterpenoids		0.114
67.556	octadecanal	638-66-4	0.096
74.551	osthol	484-12-8	0.367
80.543	isomerazin	1088-17-1	0.112
84.671	unknown (MW = 298)		0.07
		TOTAL	**99.28**

Example 5: N-terminal anchor engineering

To optimize membrane interaction of the initial *Sr*KO variants (referred to in these examples as Valencene Oxidase 1, or VO1), *E. coli* proteins anchored in the inner membrane with a cytoplasmic C-terminus were identified. An N-terminal sequence of *E. coli* yhcB was selected, which provides a single-pass transmembrane domain. 20-24 amino acids from the N-terminus of yhcB was exchanged for the original membrane anchor sequence MALLLAVF (SEQ ID NO:112), and the size of the *Sr*KO N-terminal truncation was varied from 28 to 32. See **Figure 9**. VO1 was expressed under control of a T7 promoter on a p5 plasmid. SrCPR was expressed independently from the chromosome. Strains were cultured in 96 deepwell plates at 30°C for 48 hours, in R-medium plus glycerol and dodecane overlay as already described.

As shown in **Figure 10**, n20yhcB_t29VO1 exhibited 1.2-fold productivity in total oxygenated titer compared to the average of controls. N20yhcB_t29VO1 exhibited a total oxygentated titer approximately 1.8 fold of the original 8RP anchor (not shown).

Example 6: Mutational analysis of VO1

Mutational analysis of VO1 was conducted in an effort to increase oxygenated titers. Strain MB2509 (MP6-MEP MP1-ScFPPS Fab46-VS2 MP6-ScCPR) was used as the background, which when transformed with a p5-T7-yhcB-VO1 plasmid produces about 18% nootkatone. Strains were evaluated for higher production of nootkatone.

Guided by the homology model based on P450 17A1 (Example 3) site-saturation mutagenesis of the VO active site was conducted at 18 positions, and 5

paired position libraries were constructed. First shell residues were identified through substrate docking, and non-conserved first shell residues were selected based on relative proximity and position for altering the binding pocket geometry. Paired position libraries were constructed by overlap extension PCR and Gibson assembly.

Table 12: Paired Position Libraries (numbered according to SEQ ID NO:37)

Library	**Pos. 1**	**Allowed AA**	**Pos. 2**	**Allowed AA**
1	V387	F,L,I,S,P,T,A,M	P388	S,T,A
2	M132	F,L,I,V,S,P,T,A	V133	F,L,I,S,P,T,A,M
3	L123	F,I,V,S,T,A,P,M	L126	F,I,V,S,T,A,P,M
4	V387	F,L,I,S,P,T,A,M	I322	F,L,V,S,P,A,M,T
5	I322	F,L,V,S,P,A,M,T	V133	F,L,I,S,P,T,A,M

Strains were evaluated as in Example 4 for total oxygenation of valencene. Strains were evaluated at 30°C and 22°C.

Primary screening of paired position libraries revealed that many of the variants lost activity. Library 3 contained variants with improved activity at 22°C but not 30°C. Thus, introducing two or more mutations simultaneously in the first shell residues can be determimental to activity.

Table 13: The following single position SSM was conducted (numbered according to SEQ ID NO:37)

Residue	Location	
I390	Channel	
L392	Channel	
V387	1st Shell	

E323	1st Shell	I helix
I322	1st Shell	I helix
T499	1st Shell	
Q500	1st Shell	
L231	1st Shell	F helix
L123	1st Shell	B-C loop
L126	1st Shell	B-C loop
V125	1st Shell	B-C loop
V133	1st Shell	F87 on BM3
T131	Channel	
M135	1st Shell	
L234	1st Shell	F helix
P238	1st Shell	F-G loop
M132	1st Shell	B-C loop
P388	1st Shell	

Several variants improved oxygenated titers up to 1.7-fold. Mutations at positions E323, I390, and Q500 showed several hits with improved oxygenation titer, and these positions were selected for secondary screening.

Next, back-to-consensus mutations (19 mutants) were screened in the VO1 background. Using the screening process described in Example 3, the following mutations were screened: A2T, I389L, I389V, I389A, M94V, T488D, E491K, E52A,

H46R, D191N, L150M, I495V, T468I, K344D, Q268T, R351Q, R76K, V400Q, and I444A (numbered according to SEQ ID NO:37). As shown in **Figure 13A**, more than 50% of the mutations resulted in 1.2 to 1.45 times oxygenated titers (shown as mg/L), without dramatic shifts in product profile. Improvements were seen with A2T, M94V, T488D, E52A, H46R, L150M, T468I, K344D, Q268T, R351Q, R76K, V400Q, and I444A, which were selected for secondary screening. **Figure 13B** shows the same screen plotted versus fold total oxygenated product change.

Lead variants from active site SSM (L231M, I390L, I390M, T131K, and T131Q), the N-terminal anchor variant n20yhcB_t29VOR1, and back-to-consensus mutagenesis were selected, and re-screened. The results of this secondary screen are shown in **Figure 14**. Several mutations showed a 1.1-1.4-fold improvement in oxygenated titers. To narrow the list of mutations for recombination, the same mutations were screened at 33° C to differentiate stabilizing mutations which could enable a process shift to higher temperature. As shown in **Figure 15**, six mutations (M94V, L150M, T468I, R76K, I390L, and T131Q) maintained improved productivities at 33°C. These six mutations, in addition to the lead N-terminal anchor, were selected for recombination.

Example 7: SrKO recombination library screening

The seven mutations selected after secondary screening (Example 6) were randomly incorporated into a VO recombination library by allowing either the variant or wild type at each site. The background strain was MB2509 (EGV G2 MP6-CPR) + pBAC-T7-BCD7-yhcB-VO.

Primary screening at 30°C (using the same process described in Example 5) identified several variants with up to 1.35-fold improvement in oxygenated product titers, compared to VO1. Further, select variants showed a shift in production to nootkatone, suggesting higher P450 activity (since production of nootkatone requires two oxygenation cycles). Results of primary screening are shown in **Figure 16A** (strain versus titer in mg/L). **Figure 16B** presents the same screen shown based on oxygenation capacity (total of nootkatone and nootkatol).

The recombination variants were then screened at 34°C and 37°C to select leads with improved activity and stability at higher temperature. The results of the secondary screen are shown in **Figure 17**. While the control was almost completely inactive at 37°C, six leads showed promising activity at the higher temperatures, and were selected for further screening (c11(8), b4(7), c6(1), c12(3), b6(2), and c9(6)). Based on this further screening (**Figure 18**) c6(1) was selected as the best variant based on oxygenation capacity. The six leads contain the following sets of mutations.

Table 14: Sets of mutations in lead variants from recombination library

	N20_t29yhcB	R76K	M94V	T131Q	L150M	I390L	T468I	A39H	Others
c9(6)			X			X	X		V146L
c12(3)		X		X		X	X		
b4(7)			X	X		X	X	X	
c11(8)	X	X		X		X	X	X	
c6(1)		X	X	X		X	X		
b6(2)		X	X				X		

Figure 23 (**A** and **B**) shows alignments of several engineered valencene oxidase (VO) variants as described herein, and highlights select mutations evaluated in the screening process. In Figure 23A: 8rp-t20SrKO (SEQ ID NO:106) is the SrKO sequence with a 20-amino acid truncation at the N-terminus, and the addition of an 8-amino acid membrane anchor. 8rp-t20VO0 (SEQ ID NO:107) has a truncation of 20 amino acids of the SrKO N-terminus, the addition of an 8-amino acid N-terminal anchor, and a single mutation at position 499 (numbered according to wild-type SrKO). n22yhcB-t30VO1 (SEQ ID NO:104) has a 30-amino acid truncation of the SrKO N-terminus, a membrane anchor based on 22 amino acids from *E. coli* yhcB, and eight point mutations at positions 46, 231, 284, 383, 400, 488, and 499 (with respect to SrKO wild-type). n22yhcB-t30VO2 (SEQ ID NO:105) has a 30-amino acid truncation of the

SrKO N-terminus, a membrane anchor based on 22 amino acids from *E. coli* yhcB, and nine point mutations at positions 76, 94, 131, 231, 284, 383, 390, 468, and 499 (with respect to SrKO wild-type). In Figure 24B, point mutations in VO0 (SEQ ID NO:109), VO1 (SEQ ID NO:110), and VO2 (SEQ ID NO:111) are shown against wild-type SrKO (SEQ ID NO:108) (all shown with the wild-type SrKO N-terminus for convenience).

Example 8: Cytochrome P450 Reductase screening

A set of cytochrome P450 reductases were screened for improved activity with VO1. This example was done using the strain MB2459 as the background, with pBAC-T7-BCD7-VO1(I382L)-T7BCDx-CPRx. BCD stands for BiCistronic Design, and is described in Mutalik et. al. *Nature Methods* 2013(10)4:354. Lower BCD numbers refer to higher translation rate. CPRs included SrCPR (SEQ ID NO:62), SrCPR3 (SEQ ID NO: 80), AaCPR (SEQ ID NO: 68), PgCPR (SEQ ID NO: 82), AtCPR2 (SEQ ID NO: 72), AtCPR1 (SEQ ID NO: 70), eSrCPR1 (SEQ ID NO: 76), and eATR2 (SEQ ID NO: 74). Strains were tested as in Example 5, at 30°C.

As shown in **Figure 20**, SrCPR3, which was obtained through RNA sequencing studies, exhibited a 1.3-fold improvement in oxygenated titer.

The CPR orthologs were retested at 34°C. The results are shown in **Figure 20**. Both SrCPR3 (SEQ ID NO: 80) and AaCPR (SEQ ID NO: 68) exhibited a 1.3-fold improvement in oxygenated titer, even at the higher temperature. Oxygenated titers are comparable to those obtained at 30°C.

Example 9: Alcohol Dehydrogenase enzymes to alter product profile

The ability of alcohol dehydrogenases to convert nootkatols to nootkatone was evaluated. The following ADH enzymes were evaluated:

Table 15: CPR enzymes

Gene	UniProtID	Organism
reCDH	Q9RA05	Rhodococcus erythropolis
csDH1	A0A067H4B8	Citrus sinensis
csDH2	A0A067H4S0	Citrus sinensis
csDH3		Citrus sinensis
vvDH	F6GX78	Vitis vinifera
voDH1		
csABA2	A0A067DRA0	Citrus sinensis
csDH	A0A0A0KNF1	Cucumis sativus
bdDH	I1GLS4	Brachypodium distachyon
zzSDR	F1SWA0	Zingiber zerumbet

Strains were evaluated as in Example 5, using MB2490 as the background strain (MP6-MEP FAB46-ScFPPS-L-VS1 MP6-VO1-o-SrCPR + p5-T7-BCD14-ADH). Briefly, MP6, Fab46 and T7 refer to the promoter for the attached gene or operon. Here MEP is an operon overexpressing *E. coli* dxs, idi, and ispDF genes. The L between ScFPPS and VS1 refers to a short polypeptide linker encoding (GSTGS) while -o- between VO1 and SrCPR refers to an operonic construction in which an RBS sequence is inserted between the two genes. The plus denotes a plasmid following which is described as a p5 (five copy) plasmid with a promoter, BCD (described above) and the ADH in question.

Four orthologs were identified (vvDH, csABA2, bdDH, and zzSDR) that convert nootkatol to nootkatone, resulting in more than a 3-fold increase in nootkatone titers. **Figure 21**.

REFERENCES

1. Qualley A, Dudareva N. Plant Volatiles. Encycl. Life Sci. 2010;1–9.
2. Ajikumar P, Tyo K, Carlsen S. Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. Mol. Pharm. [Internet]. 2008 [cited 2013 May 16];5(2):167–90.
3. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science [Internet]. 2010 Oct 1 [cited 2013 May 22];330(6000):70–4.
4. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature [Internet]. 2006 Apr 13 [cited 2013 Mar 1];440(7086):940–3.
5. Sevrioukova IF, Li H, Zhang H, Peterson J a, Poulos TL. Structure of a cytochrome P450-redox partner electron-transfer complex. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 1999 Mar 2;96(5):1863–8.
6. Sevrioukova IF, Poulos TL. Structural biology of redox partner interactions in P450cam monooxygenase: a fresh look at an old system. Arch. Biochem. Biophys. [Internet]. Elsevier Inc.; 2011 Mar 1 [cited 2013 Mar 26];507(1):66–74.
7. Ekroos M, Sjögren T. Structural basis for ligand promiscuity in cytochrome P450 3A4. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2006 Sep 12;103(37):13682–7.
8. Takahashi S, Yeo Y-S, Zhao Y, O'Maille PE, Greenhagen BT, Noel JP, et al. Functional characterization of premnaspirodiene oxygenase, a cytochrome P450 catalyzing regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates. J. Biol. Chem. [Internet]. 2007 Oct 26 [cited 2013 Mar 10];282(43):31744–54.
9. Morrone D, Chen X, Coates RM, Peters RJ. Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis. Biochem. J. [Internet]. 2010 Nov 1 [cited 2013 Feb 6];431(3):337–44.
10. Zhang Z, Sibbesen O. The substrate specificity of cytochrome P450 cam. Bioorganic Med. … [Internet]. 1998 [cited 2013 Sep 24];6:1501–8.
11. Stjernschantz E, van Vugt-Lussenburg BM a, Bonifacio A, de Beer SB a, van der Zwan G, Gooijer C, et al. Structural rationalization of novel drug metabolizing mutants of cytochrome P450 BM3. Proteins [Internet]. 2008 Apr [cited 2013 May 16];71(1):336–52.
12. Chen MMY, Snow CD, Vizcarra CL, Mayo SL, Arnold FH. Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. Protein Eng. Des. Sel. [Internet]. 2012 Apr [cited 2013 Mar 11];25(4):171–8.
13. Harford-Cross CF, Carmichael a B, Allan FK, England P a, Rouch D a, Wong LL. Protein engineering of cytochrome p450(cam) (CYP101) for the oxidation of polycyclic aromatic hydrocarbons. Protein Eng. [Internet]. 2000 Feb;13(2):121–8.

14. Bell SG, Chen X, Sowden RJ, Xu F, Williams JN, Wong L, et al. Molecular Recognition in (+)-r-Pinene Oxidation by Cytochrome P450cam. J. Am. Chem. Soc. 2003;125:705–14.
15. Sowden R, Yasmin S, Rees N, Bell SG, Wong L-L. Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3. Org. Biomol. Chem. [Internet]. 2005 [cited 2013 May 16];3:57–64.
16. Brandle JE, Richman A, Swanson AK, Chapman BP. Leaf Ests from Stevia rebaudiana: a resource for gene discovery in diterpene synthesis. Plant Mol. Biol. [Internet]. 2002 Nov [cited 2013 Feb 12];50(4-5):613–22.
17. Bar-even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The Moderately Efficient Enzyme : Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. Biochemistry. 2011;
18. Pleiss J. Protein design in metabolic engineering and synthetic biology. Curr. Opin. Biotechnol. [Internet]. 2011 Oct [cited 2013 Mar 5];22(5):611–7.
19. Lehmann M, Pasamontes L, Lassen SF, Wyss M. The consensus concept for thermostability engineering of proteins. Biochim. Biophys. Acta [Internet]. 2000 Dec 29 [cited 2013 Oct 17];1543(2):408–15.
20. Vazquez-Figueroa E, Yeh V, Broering JM, Chaparro-Riggers JF, Bommarius AS. Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media. Protein Eng. Des. Sel. [Internet]. 2008 Nov [cited 2013 Oct 17];21(11):673–80.
21. Dai M, Fisher HE, Temirov J, Kiss C, Phipps ME, Pavlik P, et al. The creation of a novel fluorescent protein by guided consensus engineering. Protein Eng. Des. Sel. [Internet]. 2007 Feb [cited 2013 Sep 19];20(2):69–79.
22. Fraatz M a., Riemer SJL, Stöber R, Kaspera R, Nimtz M, Berger RG, et al. A novel oxygenase from Pleurotus sapidus transforms valencene to nootkatone. J. Mol. Catal. B Enzym. [Internet]. 2009 Dec [cited 2013 Apr 17];61(3-4):202–7.
23. Krügener S, Krings U, Zorn H, Berger RG. A dioxygenase of Pleurotus sapidus transforms (+)-valencene regio-specifically to (+)-nootkatone via a stereo-specific allylic hydroperoxidation. Bioresour. Technol. [Internet]. Elsevier Ltd; 2010 Jan [cited 2013 Apr 12];101(2):457–62.
24. Zorn H, Fraatz MA, Riemer SJL, Takenberg M. Enzymatic synthesis of nootkatone. GERMANY; 2010.
25. Kaspera R, Krings U, Nanzad T, Berger RG. Bioconversion of (+)-valencene in submerged cultures of the ascomycete Chaetomium globosum. Appl. Microbiol. Biotechnol. [Internet]. 2005 Jun [cited 2013 May 16];67(4):477–83.
26. Cankar K, van Houwelingen A, Bosch D, Sonke T, Bouwmeester H, Beekwilder J. A chicory cytochrome P450 mono-oxygenase CYP71AV8 for the oxidation of (+)-valencene. FEBS Lett. [Internet]. Federation of European Biochemical Societies; 2011 Jan 3 [cited 2013 Mar 6];585(1):178–82.
27. Girhard M, Machida K, Itoh M, Schmid RD, Arisawa A, Urlacher VB. Regioselective biooxidation of (+)-valencene by recombinant E. coli expressing CYP109B1 from Bacillus subtilis in a two-liquid-phase system. Microb. Cell Fact. [Internet]. 2009 Jan [cited 2013 May 16];8(4):36.
28. Bm- P, Sowden RJ, Yasmin S, Rees NH, Bell SG, Wong L. Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450. 2005;57–64.

29. Trott O, Olson A. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. [Internet]. 2010 [cited 2013 Jul 26];31(2):455–61.l

CLAIMS:

1. A method for making an oxygenated product of a sesquiterpene, comprising, contacting the sesquiterpene with *Stevia rebaudiana* Kaurene Oxidase (*Sr*KO) or derivative thereof having sesquiterpene oxidizing activity, and recovering the oxygenated product, wherein the SrKO derivative comprises an amino acid sequence having at least 70% identity to an amino acid sequence selected from SEQ ID NO: 37, 38, or 55.

2. The method of claim 1, wherein the *Sr*KO derivative comprises an amino acid sequence having from 1 to 50 mutations independently selected from substitutions, deletions, or insertions relative to *Sr*KO (SEQ ID NO: 37, 38, or 55).

3. The method of any one of claims 1 to 2, wherein the *Sr*KO derivative comprises an amino acid sequence having at least 90% sequence identity to *Sr*KO (SEQ ID NO: 37, 38, or 55), and has valencene oxidase activity.

4. The method of any one of claims 2 to 3, wherein one or more mutations relative to *Sr*KO (SEQ ID NO: 37, 38, or 55) are selected from Table 2 or Table 6.

5. The method of claim 2, wherein the *Sr*KO is a derivative having a deletion of at least a portion of its N-terminal transmembrane region, and the addition of an inner membrane transmembrane domain from *E. coli* yhcB or derivative thereof.

6. The method of claim 5, wherein the SrKO is a derivative having a deletion of from 15 to 35 amino acids of its N-terminal transmembrane domain, and the addition of from 15 to 25 amino acids of the transmembrane domain from *E. coli* yhcB or derivative thereof.

7. The method of claim 6, wherein the N-terminal transmembrane domain from yhcB comprises the amino acid sequence of SEQ ID NO: 118, or an amino acid sequence having from 1 to 5 amino acid mutations with respect to SEQ ID NO: 118.

8. The method of any one of claims 1 to 7, wherein the SrKO is a derivative having one or more mutations at positions selected from 46, 76, 94, 131, 231, 284, 383, 390, 400, 444, 468, 488 and 499 relative to SEQ ID NO:37.

9. The method of claim 8, wherein the SrKO is a derivative having one or more mutations selected from R76K, M94V, T131Q, F231L, H284Q, R383K, I390L, T468I, and T499N relative to SEQ ID NO:37.

10. The method of claim 1, wherein the SrKO is a derivative comprising an amino acid sequence selected from SEQ ID NOS:55-61, 104, and 105 or an amino acid sequence having from one to twenty mutations relative to a sequence selected from SEQ ID NOS: 55-61, 104 and 105.

11. The method of any one of claims 1 to 10, wherein the sesquiterpene is valencene, germacrene (A, B, C, D, or E), farnesene, farnesol, patchoulol, cadinene, cedrol, humulene, longifolene, and/or bergamotene, β-Ylangene, β-santalol, β-santalene, α-santalene, α-santalol, ß-vetivone, a-vetivone, khusimol, bisabolene, β-aryophyllene, Longifolene; α-sinensal; α-bisabolol, (-)-β-Copaene, (-)-α-Copaene, 4(Z),7(Z)-ecadienal, cedrol, cedrene, cedrol, guaiol, (-)-6,9-guaiadiene, bulnesol, guaiol, ledene, ledol, lindestrene, and alpha-bergamotene.

12. The method of claim 1 to 11, wherein the predominant sesquiterpene is valencene, germacrene (A, B, C, D, or E), farnesene, farnesol, patchoulol, cadinene, cedrol, humulene, longifolene, and/or bergamotene, β-Ylangene, β-santalol, β-santalene, α-santalene, α-santalol, ß-vetivone, a-vetivone, khusimol, bisabolene, β-aryophyllene, Longifolene; α-sinensal; α-bisabolol, (-)-β-Copaene, (-)-α-Copaene, 4(Z),7(Z)-ecadienal, cedrol, cedrene, cedrol, guaiol, (-)-6,9-guaiadiene, bulnesol, guaiol, ledene, ledol, lindestrene, and alpha-bergamotene.

13. The method of claim 12, wherein the predominant sesquiterpene is valencene.

14. The method of claim 13, wherein the predominant oxygenated product is nootkatone.

15. The method of any one of claims 1 to 14, wherein the contacting takes place in a host cell, which is a bacterium.

16. The method of claim 15, wherein the host cell is *E. coli*, *Bacillus subtillus*, or *Pseudomonas putida*.

17. The method of any one of claims 1 to 14, wherein the contacting takes place in a host cell which is a yeast.

18. The method of claim 17, wherein the host cell is a species of *Saccharomyces*, *Pichia*, or *Yarrowia*, including *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Yarrowia lipolytica*.

19. The method of any one of claims 15 to 18, wherein isopentyl pyrophosphate (IPP) is produced by the host cell by metabolic flux through an endogenous or heterologous MEP or MVA pathway.

20. The method of claim 19, wherein the sesquiterpene is produced at least in part by metabolic flux through an MEP pathway, and wherein the host cell has at least one additional copy of a dxs, ispD, ispF, and/or idi gene.

21. The method of any one of claims 15 to 20, wherein the host cell further expresses a farnesyl pyrophosphate synthase.

22. The method of any one of claims 15 to 21, wherein the host cell further expresses a heterologous sesquiterpene synthase.

23. The method of claim 22, wherein the sesquiterpene synthase is a valencene synthase.

24. The method of claim 23, wherein the valencene synthase is *Citrus sinensis* valencene synthase (CsVS) or *Vitis vinfera* valencene synthase (VvVS) or derivative thereof.

25. The method of claim 24, wherein the VvVS comprises an amino acid sequence having from 1 to 40 mutations independently selected from substitutions, deletions, or insertions with respect to VvVS (SEQ ID NO:1).

26. The method of claim 24, wherein the CsVS comprises an amino acid sequence having from 1 to 40 mutations independently selected from substitutions, deletions, or insertions with respect to CsVS (SEQ ID NO:12).

27. The method of any one of claims 1 to 26, wherein the SrKO is expressed in a host cell that co-expresses a cytochrome P450 reductase.

28. The method of claim 27, wherein the P450 reductase is a CPR from *Stevia sp.*

29. The method of any one of claims 1 to 28, wherein the *Sr*KO is expressed as a fusion with a cytochrome P450 reductase partner.

30. The method of any one of claims 1 to 29, wherein the host cell further expresses an alcohol dehydrogenase producing nootkatone.

31. The method of any one of claims 1 to 30, wherein oxidized oil is extracted from aqueous reaction medium followed by fractional distillation.

32. The method of claim 31, wherein the oxidized oil is extracted from aqueous reaction medium using an organic solvent.

33. The method of claim 31, wherein sesquiterpene and sesquiterpenoid components of fractions are measured quantitatively by GC/MS, followed by blending of the fractions to generate a desired nootkatone-containing ingredient.

34. The method of claim 31 or 32, wherein the nootkatone fraction is recovered.

35. An *Sr*KO derivative comprising at least one mutation with respect to SEQ ID NO: 37, 38, or 55 that increases valencene oxidase activity, wherein the SrKO derivative comprises an amino acid sequence having at least 70% identity to an amino acid sequence selected from SEQ ID NO: 37, 38, or 55.

36. The SrKO derivative of claim 35, wherein the *Sr*KO derivative has a deletion of at least a portion of its N-terminal transmembrane region, and the addition of an inner membrane transmembrane domain from *E. coli* yhcB or derivative thereof.

37. The *Sr*KO derivative of claim 36, wherein the *Sr*KO derivative has a deletion of from 15 to 35 amino acids of its N-terminal transmembrane domain, and the addition of from 15 to 25 amino acids of the transmembrane domain from *E. coli* yhcB or derivative thereof.

38. The *Sr*KO derivative of claim 37, wherein the N-terminal transmembrane domain comprises the amino acid sequence of SEQ ID NO: 118, or an amino acid sequence having from 1 to 5 amino acid mutations with respect to SEQ ID NO: 118.

39. The *Sr*KO derivative of claims 35 to 38, wherein the *Sr*KO derivative has one or more mutations at positions selected from 46, 76, 94, 131, 231, 284, 383, 390, 400, 444, 468, 488 and 499 relative to SEQ ID NO:37.

40. The *Sr*KO derivative of claim 39, wherein the *Sr*KO derivative has one or more mutations selected from R76K, M94V, L107M, T131Q, F231L, H284Q, R383K, I390L, T468I, and T499N relative to SEQ ID NO:37.

41. The *Sr*KO derivative of claim 40, wherein the *Sr*KO derivative comprises an amino acid sequence selected from SEQ ID NOS:55-61, 104, and 105.

42. A host cell producing an oxygenated sesquiterpene: wherein the host cell produces isopentyl pyrophosphate (IPP); the host cell expresses a farnesyl pyrophosphate synthase, a sesquiterpene synthase, and SrKO or derivative thereof having valencene

oxidase activity, and wherein the *Sr*KO derivative has from 1 to 40 mutations relative to SEQ ID NO: 37, 38, or 55.

43. A host cell producing an oxygenated sesquiterpene: wherein the host cell produces isopentyl pyrophosphate (IPP); the host cell expresses a farnesyl pyrophosphate synthase, a sesquiterpene synthase, and SrKO or derivative thereof having valencene oxidase activity, and wherein the *Sr*KO derivative comprises an amino acid sequence having at least 70% sequence identity to SEQ ID NO: 37, 38, or 55, and has valencene oxidase activity.

44. A host cell producing an oxygenated sesquiterpene: wherein the host cell produces isopentyl pyrophosphate (IPP); the host cell expresses a farnesyl pyrophosphate synthase, a sesquiterpene synthase, and SrKO or derivative thereof having valencene oxidase activity, and wherein the SrKO derivative has one or more mutations relative to SEQ ID NO:37 or 38 that are selected from Table 2 or Table 6.

45. The host cell of claim 44, wherein at least 2 mutations are selected from Table 2 and/or Table 6.

46. A host cell producing an oxygenated sesquiterpene: wherein the host cell produces isopentyl pyrophosphate (IPP); the host cell expresses a farnesyl pyrophosphate synthase, a sesquiterpene synthase, and SrKO or derivative thereof having valencene oxidase activity, and wherein the *Sr*KO derivative has a deletion of at least a portion of its N-terminal transmembrane region, and the addition of an inner membrane transmembrane domain from *E. coli* yhcB or derivative thereof.

47. The host cell of claim 46, wherein the SrKO derivative has a deletion of from 15 to 35 amino acids of its N-terminal transmembrane domain, and the addition of from 15 to 25 amino acids of the transmembrane domain from E. coli yhcB or derivative thereof.

48. A host cell producing an oxygenated sesquiterpene: wherein the host cell produces isopentyl pyrophosphate (IPP); the host cell expresses a farnesyl pyrophosphate synthase, a sesquiterpene synthase, and SrKO or derivative thereof having valencene

oxidase activity, wherein the *Sr*KO derivative comprises an amino acid sequence selected from SEQ ID NOS:55-61, 104, or 105; or an amino acid sequence having from one to twenty mutations relative to a sequence selected from SEQ ID NOS:55-61, 104, or 105 with the proviso that the amino acid sequence has one or more mutations at positions selected from 46, 76, 94, 131, 231, 284, 383, 390, 400, 444, 468, 488 and 499 relative to SEQ ID NO:37, or the proviso that the amino acid sequence has one or more of the mutations selected from R76K, M94V, T131Q, F231L, H284Q, R383K, I390L, T468I, and T499N relative to SEQ ID NO:37.

49. The host cell of any one of claims 42 to 48, wherein the host cell is an *E. coli* having at least one additional copy of a dxs, ispD, ispF, and/or idi gene and wherein the host cell produces nootkatone.

50. The host cell of of any one of claims 42 to 49, wherein the sesquiterpene synthase is a *Vitis vinifera* valencene synthase (VvVS) derivative having from 1 to 40 mutations independently selected from substitutions, deletions, or insertions with respect to VvVS (SEQ ID NO:1) and further one or more mutations selected from Table 3.

51. The host cell of any one of claims 42 to 50, wherein the SrKO or derivative thereof is expressed in a host cell that co-expresses a cytochrome P450 reductase and wherein the P450 reductase comprises an amino acid sequence selected from SEQ ID NOS: 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, and 82, or a derivative thereof.

52. The host cell of any one of claims 42 to 51, wherein the host cell further expresses an alcohol dehydrogenase enzyme that converts nootkatol to nootkatone.